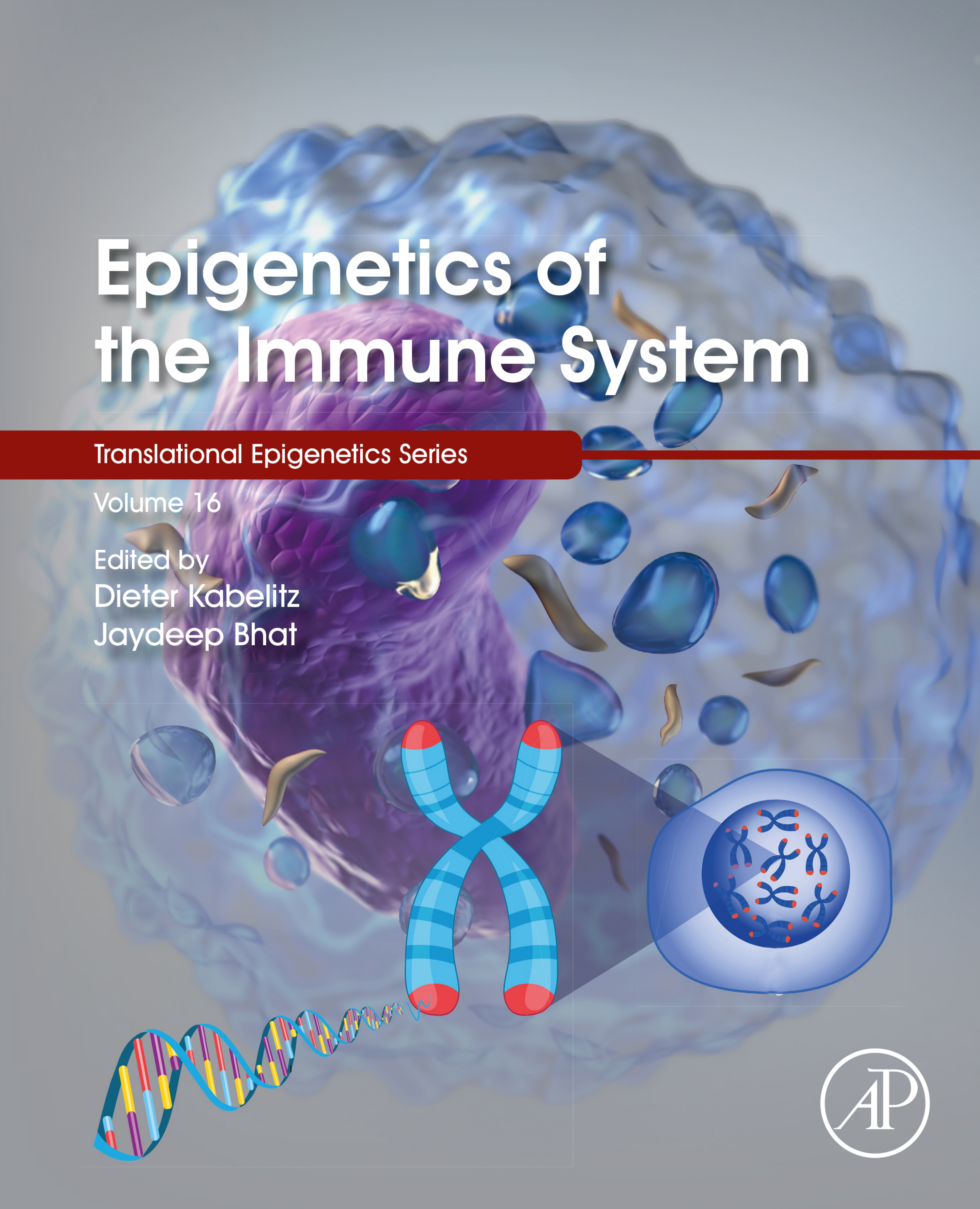


Epigenetics of the Immune System

Translational Epigenetics Series

Volume 16

Edited by
Dieter Kabelitz
Jaydeep Bhat

EPIGENETICS OF THE IMMUNE SYSTEM

Translational Epigenetics Series

Trygve O. Tollefsbol, Series Editor

Transgenerational Epigenetics
Edited by Trygve O. Tollefsbol, 2014

Personalized Epigenetics
Edited by Trygve O. Tollefsbol, 2015

Epigenetic Technological Applications
Edited by Y. George Zheng, 2015

Epigenetic Cancer Therapy
Edited by Steven G. Gray, 2015

DNA Methylation and Complex Human Disease
By Michel Neidhart, 2015

Epigenomics in Health and Disease
Edited by Mario F. Fraga and
Agustin F. F Fernández, 2015

Epigenetic Gene Expression and Regulation
Edited by Suming Huang, Michael Litt and
C. Ann Blakey, 2015

Epigenetic Biomarkers and Diagnostics
Edited by Jose Luis García-Giménez, 2015

Drug Discovery in Cancer Epigenetics
Edited by Gerda Egger and
Paola Barbara Arimondo, 2015

Medical Epigenetics
Edited by Trygve O. Tollefsbol, 2016

Chromatin Signaling and Diseases
Edited by Olivier Binda and
Martin Fernandez-Zapico, 2016

Genome Stability
Edited by Igor Kovalchuk and
Olga Kovalchuk, 2016

Chromatin Regulation and Dynamics
Edited by Anita Göndör, 2016

Neuropsychiatric Disorders and Epigenetics
Edited by Dag H. Yasui, Jacob Peedicayil and
Dennis R. Grayson, 2016

Polycomb Group Proteins
Edited by Vincenzo Pirrotta, 2016

Epigenetics and Systems Biology
Edited by Leonie Ringrose, 2017

Cancer and Noncoding RNAs
Edited by Jayprokas Chakrabarti and
Sanga Mitra, 2017

Nuclear Architecture and Dynamics
Edited by Christophe Lavelle and
Jean-Marc Victor, 2017

Epigenetic Mechanisms in Cancer
Edited by Sabita Saldanha, 2017

Epigenetics of Aging and Longevity
Edited by *Alexey Moskalev and*
Alexander M. Vaiserman, 2017

The Epigenetics of Autoimmunity
Edited by Rongxin Zhang, 2018

Epigenetics in Human Disease, Second Edition
Edited by Trygve O. Tollefsbol, 2018

Epigenetics of Chronic Pain
Edited by Guang Bai and
Ke Ren, 2018

Epigenetics of Cancer Prevention
Edited by Anupam Bishayee and
Deepak Bhatia, 2018

Computational Epigenetics and Diseases
Edited by Loo Keat Wei, 2019

Prognostic Epigenetics
Edited by Shilpy Sharma, 2019

RNA-Based Regulation in Human Health and Disease (ISBN: 9780128171936)
Edited by Pandey Rajesh, 2020

Stem Cell Epigenetics (ISBN: 9780128140857)
Edited by Eran Meshorer and
Giuseppe Testa, 2020

Translational Epigenetics

EPIGENETICS OF THE IMMUNE SYSTEM

Volume 16

Edited by

DIETER KABELITZ
Institute of Immunology, University of Kiel, Kiel, Germany

JAYDEEP BHAT
Institute of Immunology, University of Kiel, Kiel, Germany

ELSEVIER

ACADEMIC PRESS
An imprint of Elsevier

Academic Press is an imprint of Elsevier
125 London Wall, London EC2Y 5AS, United Kingdom
525 B Street, Suite 1650, San Diego, CA 92101, United States
50 Hampshire Street, 5th Floor, Cambridge, MA 02139, United States
The Boulevard, Langford Lane, Kidlington, Oxford OX5 1GB, United Kingdom

© 2020 Elsevier Inc. All rights reserved.

No part of this publication may be reproduced or transmitted in any form or by any means, electronic or mechanical, including photocopying, recording, or any information storage and retrieval system, without permission in writing from the publisher. Details on how to seek permission, further information about the Publisher's permissions policies and our arrangements with organizations such as the Copyright Clearance Center and the Copyright Licensing Agency, can be found at our website: www.elsevier.com/permissions.

This book and the individual contributions contained in it are protected under copyright by the Publisher (other than as may be noted herein).

Notices
Knowledge and best practice in this field are constantly changing. As new research and experience broaden our understanding, changes in research methods, professional practices, or medical treatment may become necessary.

Practitioners and researchers must always rely on their own experience and knowledge in evaluating and using any information, methods, compounds, or experiments described herein. In using such information or methods they should be mindful of their own safety and the safety of others, including parties for whom they have a professional responsibility.

To the fullest extent of the law, neither the Publisher nor the authors, contributors, or editors, assume any liability for any injury and/or damage to persons or property as a matter of products liability, negligence or otherwise, or from any use or operation of any methods, products, instructions, or ideas contained in the material herein.

Library of Congress Cataloging-in-Publication Data
A catalog record for this book is available from the Library of Congress

British Library Cataloguing-in-Publication Data
A catalogue record for this book is available from the British Library

ISBN 978-0-12-817964-2

For information on all Academic Press publications
visit our website at https://www.elsevier.com/books-and-journals

Publisher: Andre Gerhard Wolff
Acquisitions Editor: Linda Versteeg-buschman
Editorial Project Manager: Timothy Bennett
Production Project Manager: Maria Bernard
Cover Designer: Miles Hitchen

Typeset by SPi Global, India

ELSEVIER Book Aid International
Working together
to grow libraries in
developing countries
www.elsevier.com • www.bookaid.org

Contents

Contributors *xi*
Preface *xv*

1. An introduction to immunology and epigenetics **1**

Jaydeep Bhat, Dieter Kabelitz

Development of immune cells 2
Dynamics of the immune response 5
Immunological memory 8
Basic overview of epigenetic regulation 9
Genome architecture at primary structure scale: Cis-regulation through DNA elements 10
Genome architecture at secondary structure scale: Trans-regulation through chromatin packaging and nucleosome positioning 11
Genome architecture at tertiary structure scale: Long-range chromatic interactions and gene regulation within chromatin territory and compartments 13
Posttranscriptional regulation of RNA expression 14
Posttranslational modifications: Chromatin remodeling complexes and histones 16
Posttranslational modifications: Nonhistone proteins 18
Advancing technology and use of interdisciplinary approaches 19
References 20

2. Plant epigenetics and the 'intelligent' priming system to combat biotic stress **25**

Hemlata Kotkar, Ashok Giri

Introduction 25
Plant DNA methylation 26
De novo DNA methylation 27
Maintenance DNA methylation 28
Histone modifications 29
RNA-associated silencing 29
Epigenetics of plant microbe interactions 29
Epigenetics of plant insect interactions 30
Epigenetics of immune system and memory in plants 32
Plant epigenetics: Model plants and application in agriculture 33
Somatic embryogenesis 34
Heterosis 34

Conclusions 35
Acknowledgments 36
References 36

3. Understanding immune system development: An epigenetic perspective 39
Ayush Madhok, Anjali deSouza, Sanjeev Galande

Introduction: An outline of the chapter 39
Epigenetic modifications and their functional output 40
Epigenetic mechanisms regulating immune cell development 45
Concluding remarks and future perspective 63
Acknowledgments 63
References 64
Further reading 76

4. Epigenetic mechanisms in the regulation of lymphocyte differentiation 77
Nina Schmolka, Bruno Silva-Santos, Anita Q. Gomes

Introduction 77
Part I: Chromatin-based epigenetic mechanisms in lymphocyte differentiation 80
Part II: RNA-based mechanisms of regulation of lymphocyte differentiation 89
Concluding remarks 105
References 106
Further reading 116

5. Epigenetics mechanisms driving immune memory cell differentiation and function 117
Stephen J. Turner, Jasmine Li, Brendan E. Russ

Introduction 117
Functional heterogeneity within the memory T-cell pool 119
Histone methylation and pattern, acquisition of function 120
DNA methylation and its role in regulating gene transcription 122
Epigenetic regulation in the acquisition of lineage-specific T-cell function 123
Epigenetic regulation in the acquisition of CD8 T-cell function 124
Epigenetic mapping of the differentiation pathway that leads to T-cell memory 126
The role of CD8$^+$ T-cell-specific transcription factors in chromatin remodeling and acquisition of function 127
Active regulation of chromatin state is a key factor in CD8$^+$ T-cell effector vs memory fate decisions 129
Conclusion 132
References 133

6. Microbiota in the context of epigenetics of the immune system **139**

Katarzyna Placek

Introduction 139
Epigenetic mechanisms 141
Gut microbiome and epigenetics of immune cells 142
Skin microbiota and epigenetics of immune cells 148
Microbiome and epigenetics of nonmucosal immune cells 149
Conclusions and prospects 152
References 153

7. Sequencing technologies for epigenetics: From basics to applications **161**

Rosario Michael Piro

Introduction to high-throughput sequencing 163
Library preparation 164
Flow cell preparation 165
Sequencing by synthesis 167
Applications of sequencing technologies for epigenetics 169
Data processing and computational analysis 172
Future perspectives 178
References 178

8. Advances in single-cell epigenomics of the immune system **185**

Jonas Schulte-Schrepping, Humberto J. Ferreira, Adem Saglam, Emily Hinkley, Joachim L. Schultze

Introduction 185
Single-cell epigenomics technologies 188
Chromatin structure 193
Computational challenges and solutions 196
Studying the immune system using single-cell epigenomics 202
Conclusions and future perspectives 208
References 209

9. Machine learning and deep learning for the advancement of epigenomics **217**

Magdalena A. Machnicka, Bartek Wilczynski

The “epigenetic code” problem 217
Progress of machine learning: Classification versus non-supervised learning 218
Unsupervised approaches 221
Supervised approaches 223

Conclusion 233
Acknowledgments 234
References 234

10. Systems immunology meets epigenetics 239
Wenhui Li, Ziyi Chen, Aiping Wu, F. Xiao-Feng Qin, Lianjun Zhang

Epigenetic modifications within the immune system 239
Systems approach for deconvoluting immune cell composition 243
References 250

11. Epigenetic deregulation of immune cells in autoimmune and autoinflammatory diseases 253
Javier Rodríguez-Ubreva, Tianlu Li, Esteban Ballestar

Relevance of epigenetics for immune deregulation in autoimmune/autoinflammatory disorders 253
Epigenetic dysregulation in autoimmune diseases 254
Epigenetic dysregulation in autoinflammatory diseases 258
Epigenetic biomarkers in autoimmunity 260
Targeting epigenetic defects 262
References 263

12. Epigenetics of allergies: From birth to childhood 269
Avery DeVries, Donata Vercelli

Neonatal DNA methylation profiles as predictors of the trajectory to asthma and allergy 273
DNA methylation profiles in patients with childhood asthma and allergy 276
What have we learned so far? 279
References 281

13. Epigenetic regulation of normal hematopoiesis and its dysregulation in hematopoietic malignancies 285
Mark Hartmann, Mariam Hakobyan, Jens Langstein, Maximilian Schönung, Sina Stäble, Aurore Touzart, Justyna A. Wierzbinska, Daniel B. Lipka

Epigenetic regulation of normal hematopoiesis 286
Mutations of epigenetic regulators in clonal hematopoiesis and in hematopoietic neoplasms 290
Epigenetic deregulation in T-cell acute lymphoblastic leukemia 297
Cell-of-origin identification using epigenetics: Chronic lymphocytic leukemia as a model 300

Epigenomics for disease classification and as a diagnostic tool 304
Epigenetic therapies in hematopoietic neoplasms 307
Acknowledgment 310
References 310

14. Impact of epigenetic modifiers on the immune system 315

Sajad Ahmad Bhat, Shalini Kashipathi Sureshbabu, Chinna Susan Philip, Shubhada Chiplunkar

Overview 315
Epigenetic modifiers 316
Immunomodulatory effects of epigenetic modifiers 319
Clinical significance of epigenetic modifiers 325
Combination therapy involving epigenetic modifiers 342
Conclusion 345
Acknowledgments 346
References 346

Index 353

This page intentionally left blank

Contributors

Esteban Ballestar
Epigenetics and Immune Disease Group, Josep Carreras Research Institute (IJC), Barcelona, Spain

Jaydeep Bhat
Institute of Immunology, University of Kiel, Kiel, Germany

Sajad Ahmad Bhat
Chiplunkar Laboratory, Advanced Centre for Treatment, Research and Education in Cancer (ACTREC), Tata Memorial Centre; Homi Bhabha National Institute, Mumbai, India

Ziyi Chen
Suzhou Institute of Systems Medicine, Suzhou, Jiangsu; Center for Systems Medicine, Institute of Basic Medical Sciences, Chinese Academy of Medical Sciences and Peking Union Medical College, Beijing, China

Shubhada Chiplunkar
Chiplunkar Laboratory, Advanced Centre for Treatment, Research and Education in Cancer (ACTREC), Tata Memorial Centre; Homi Bhabha National Institute, Mumbai, India

Anjali deSouza
Laboratory of Chromatin Biology and Epigenetics, Department of Biology, Indian Institute of Science Education and Research, Pune, India

Avery DeVries
Asthma & Airway Disease Research Center, University of Arizona, Tucson, AZ, United States

Humberto J. Ferreira
Platform for Single Cell Genomics and Epigenomics at the German Center for Neurodegenerative Diseases and the University of Bonn, Bonn, Germany

Sanjeev Galande
Laboratory of Chromatin Biology and Epigenetics, Department of Biology, Indian Institute of Science Education and Research, Pune, India

Ashok Giri
Biochemical Sciences Division, CSIR-National Chemical Laboratory, Pune, India

Anita Q. Gomes
Instituto de Medicina Molecular João Lobo Antunes, Faculdade de Medicina, Universidade de Lisboa, Lisboa; H&TRC Health & Technology Research Center, ESTeSL—Escola Superior de Tecnologia da Saúde, Instituto Politécnico de Lisboa, Lisbon, Portugal

Mariam Hakobyan
Section Translational Cancer Epigenomics, Division of Translational Medical Oncology, National Center of Tumor Diseases (NCT) & German Cancer Research Center (DKFZ); Faculty of Biosciences, Heidelberg University, Heidelberg, Germany

Mark Hartmann
Section Translational Cancer Epigenomics, Division of Translational Medical Oncology, National Center of Tumor Diseases (NCT) & German Cancer Research Center (DKFZ), Heidelberg, Germany

Emily Hinkley
Platform for Single Cell Genomics and Epigenomics at the German Center for Neurodegenerative Diseases and the University of Bonn, Bonn, Germany

Dieter Kabelitz
Institute of Immunology, University of Kiel, Kiel, Germany

Hemlata Kotkar
Department of Botany, Savitribai Phule Pune University, Pune, India

Jens Langstein
Section Translational Cancer Epigenomics, Division of Translational Medical Oncology, National Center of Tumor Diseases (NCT) & German Cancer Research Center (DKFZ); Faculty of Biosciences, Heidelberg University, Heidelberg, Germany

Jasmine Li
Department of Microbiology, Biomedical Discovery Institute, Monash University, Clayton, VIC, Australia

Tianlu Li
Epigenetics and Immune Disease Group, Josep Carreras Research Institute (IJC), Barcelona, Spain

Wenhui Li
Suzhou Institute of Systems Medicine, Suzhou, Jiangsu; Center for Systems Medicine, Institute of Basic Medical Sciences, Chinese Academy of Medical Sciences and Peking Union Medical College, Beijing, China

Daniel B. Lipka
Section Translational Cancer Epigenomics, Division of Translational Medical Oncology, National Center of Tumor Diseases (NCT) & German Cancer Research Center (DKFZ), Heidelberg, Germany

Magdalena A. Machnicka
Institute of Informatics, Faculty of Mathematics, Informatics and Mechanics, University of Warsaw, Poland

Ayush Madhok
Laboratory of Chromatin Biology and Epigenetics, Department of Biology, Indian Institute of Science Education and Research, Pune, India

Chinna Susan Philip
Chiplunkar Laboratory, Advanced Centre for Treatment, Research and Education in Cancer (ACTREC), Tata Memorial Centre; Homi Bhabha National Institute, Mumbai, India

Rosario Michael Piro*
Department of Mathematics and Computer Science, Freie Universität Berlin; Institute of Medical Genetics and Human Genetics, Charité-Universitätsmedizin Berlin, Berlin; German Cancer Consortium (DKTK) partner site Berlin, and German Cancer Research Center (DKFZ), Heidelberg, Germany

Katarzyna Placek
Molecular Immunology and Cell Biology Unit, Life and Medical Sciences Institute, University of Bonn, Bonn, Germany

F. Xiao-Feng Qin
Suzhou Institute of Systems Medicine, Suzhou, Jiangsu; Center for Systems Medicine, Institute of Basic Medical Sciences, Chinese Academy of Medical Sciences and Peking Union Medical College, Beijing, China

Javier Rodríguez-Ubreva
Epigenetics and Immune Disease Group, Josep Carreras Research Institute (IJC), Barcelona, Spain

Brendan E. Russ
Department of Microbiology, Biomedical Discovery Institute, Monash University, Clayton, VIC, Australia

Adem Saglam
Platform for Single Cell Genomics and Epigenomics at the German Center for Neurodegenerative Diseases and the University of Bonn, Bonn, Germany

Nina Schmolka
Department of Molecular Mechanisms of Disease, University of Zurich, Zurich, Switzerland

Maximilian Schönung
Section Translational Cancer Epigenomics, Division of Translational Medical Oncology, National Center of Tumor Diseases (NCT) & German Cancer Research Center (DKFZ); Faculty of Biosciences, Heidelberg University, Heidelberg, Germany

Jonas Schulte-Schrepping
Genomics and Immunoregulation, Life & Medical Sciences (LIMES) Institute, University of Bonn, Bonn, Germany

---

*Current address: Department of Electronics, Informatics and Bioengineering (DEIB), Polytechnic University of Milan, Milan, Italy

Joachim L. Schultze
Genomics and Immunoregulation, Life & Medical Sciences (LIMES) Institute, University of Bonn; Platform for Single Cell Genomics and Epigenomics at the German Center for Neurodegenerative Diseases and the University of Bonn, Bonn, Germany

Bruno Silva-Santos
Instituto de Medicina Molecular João Lobo Antunes, Faculdade de Medicina, Universidade de Lisboa, Lisboa, Portugal

Sina Stäble
Section Translational Cancer Epigenomics, Division of Translational Medical Oncology, National Center of Tumor Diseases (NCT) & German Cancer Research Center (DKFZ); Faculty of Biosciences, Heidelberg University, Heidelberg, Germany

Shalini Kashipathi Sureshbabu
Chiplunkar Laboratory, Advanced Centre for Treatment, Research and Education in Cancer (ACTREC), Tata Memorial Centre; Homi Bhabha National Institute, Mumbai, India

Aurore Touzart
Division of Cancer Epigenomics, German Cancer Research Center (DKFZ), Heidelberg, Germany

Stephen J. Turner
Department of Microbiology, Biomedical Discovery Institute, Monash University, Clayton; Department of Microbiology and Immunology, The Peter Doherty Institute for Infection and Immunity, The University of Melbourne, Parkville, VIC, Australia

Donata Vercelli
Asthma & Airway Disease Research Center; Department of Cellular and Molecular Medicine; Arizona Center for the Biology of Complex Diseases, University of Arizona, Tucson, AZ, United States

Justyna A. Wierzbinska
Faculty of Biosciences, Heidelberg University; Division of Cancer Epigenomics, German Cancer Research Center (DKFZ), Heidelberg, Germany

Bartek Wilczynski
Institute of Informatics, Faculty of Mathematics, Informatics and Mechanics, University of Warsaw, Poland

Aiping Wu
Suzhou Institute of Systems Medicine, Suzhou, Jiangsu; Center for Systems Medicine, Institute of Basic Medical Sciences, Chinese Academy of Medical Sciences and Peking Union Medical College, Beijing, China

Lianjun Zhang
Suzhou Institute of Systems Medicine, Suzhou, Jiangsu; Center for Systems Medicine, Institute of Basic Medical Sciences, Chinese Academy of Medical Sciences and Peking Union Medical College, Beijing, China

Preface

Immunology has developed at an unprecedented pace in recent years. Major breakthroughs in basic immunology in the last century, like the discovery of antibody diversity, natural killer cells, T-cell receptors, monoclonal antibodies, interleukins, to name but a few, have been followed by similar groundbreaking progress in translational and clinical immunology, like the use of monoclonal antibodies to treat certain types of cancer, or to give exhausted T cells a break. A key feature of the immune system is its enormous plasticity. All immune cells spanning from innate granulocytes and macrophages to the sophisticated adaptive B and T cells expressing seemingly unlimited antibody- and T-cell receptor repertoires can arise from one single common hematopoietic progenitor cell. And peripheral $CD4^+$ T cells are not a homogeneous cell population; depending on the surrounding micromilieu, they can differentiate into functionally diverse effector cells expressing distinct transcription factors and cytokines, such as Th1, Th2, Th9, Th17, and others. All differentiation processes in the immune system are heavily modulated and regulated by epigenetic mechanisms including DNA (de)methylation and histone modifications. Epigenetics describes heritable alterations which are not fixed in the genome. Today, it is clear that almost all (if not all) biological processes are impacted by epigenetic regulation, but this is perhaps most evident (and important) in the immune system. Therefore, it is a timely subject to analyze in some detail how the immune cell development and the functional differentiation of mature T cells are regulated by epigenetics, and how immunological diseases like autoimmunity, allergies, or lymphoma development are affected by epigenetic regulation. The analysis of epigenetic regulation requires latest technologies in sequencing, single-cell analysis, and bioinformatic tools including machine learning, as well as systems immunology approaches.

The goal of this book is to give an overview on basic principles of epigenetic regulation and how this impacts selected aspects of the immune system. Furthermore, the book provides an excellent overview of the most advanced and important technologies for epigenetic analysis. The chapters on selected clinical topics give important insights into how certain diseases are modulated by epigenetic processes. Last but not least, therapeutic perspectives based on epigenetic modulation are also discussed. Certain drugs affecting DNA methylation or histone modification are already in the clinics—and more are on the horizon.

This book shall be of interest to a wide audience. We believe that immunologists with limited background in epigenetics will greatly benefit from studying the pathways through which epigenetics regulates immune cell differentiation. On the other hand, experts in epigenetics with limited experience in immunology will appreciate to find out that the immune system is an excellent biological system to study epigenetics. Finally, clinicians treating patients with immune-related disorders will realize the impact of epigenetics on (some) diseases, as well as epigenetic mechanisms as a potential target for novel therapies.

We are very grateful to all the colleagues who contributed a chapter to this book. Without your willingness to share your expertise, it would not have been possible to present this excellent collection of insightful chapters on the fascinating aspects of the role of epigenetics in immunology—from basic to clinical immunology and translational/therapeutic perspectives. We also acknowledge those colleagues who submitted their contribution early and on time to meet the original deadline—a multiauthor endeavor like this book is only finished once the last chapter is submitted.

Last but not least, we are very grateful to Timothy J. Bennett, Editorial Project Manager at Elsevier Publisher, for his excellent guidance through this project. It has been a pleasure to work with him!

Dieter Kabelitz, Jaydeep Bhat
Institute of Immunology, University of Kiel, Kiel, Germany

CHAPTER 1

An introduction to immunology and epigenetics

Jaydeep Bhat, Dieter Kabelitz
Institute of Immunology, University of Kiel, Kiel, Germany

Contents

Development of immune cells 2
Dynamics of the immune response 5
Immunological memory 8
Basic overview of epigenetic regulation 9
Genome architecture at primary structure scale: Cis-regulation through DNA elements 10
Genome architecture at secondary structure scale: Trans-regulation through chromatin packaging and nucleosome positioning 11
Genome architecture at tertiary structure scale: Long-range chromatic interactions and gene regulation within chromatin territory and compartments 13
Posttranscriptional regulation of RNA expression 14
Posttranslational modifications: Chromatin remodeling complexes and histones 16
Posttranslational modifications: Nonhistone proteins 18
Advancing technology and use of interdisciplinary approaches 19
References 20

The immune system not only protects the organism against danger from the outside world (e.g., infections caused by bacteria, viruses, fungi, or parasites), but also maintains tolerance to self, and monitors cellular integrity by sensing stressed and malignant cells ("immune surveillance"). A key feature of the immune system is immunological memory, i.e., the capacity to rapidly respond with increased intensity upon secondary challenge with the same antigen (which forms the basis of vaccination). The immune system is composed of a multitude of cells and soluble factors that are important for the orchestration of immune responses and the bidirectional interaction between the immune system and other organ systems (e.g., the nervous system). Historically, the components of the immune system have been categorized as nonspecific ("innate") and specific ("adaptive") branches. The adaptive arm of immunity comprises B lymphocytes (B cells), which produce antibodies (immunoglobulins, Ig), and T lymphocytes (T cells), which are specific effector cells of cellular immunity. During development, both B cells and T cells undergo a rearrangement of germline-encoded genes coding for Ig and T-cell receptor (TCR), respectively, leading to an almost unlimited antibody and TCR repertoire in mature B-cell and T-cell compartments. Cells of

Epigenetics of the Immune System
https://doi.org/10.1016/B978-0-12-817964-2.00001-0
© 2020 Elsevier Inc.
All rights reserved.

the innate immune system including granulocytes, monocytes/macrophages, innate lymphoid cells (ILCs), natural killer (NK) cells do not express antigen-specific receptors (like B and T cells), but express a plethora of activating and inhibitory receptors that govern their functional activity. The communication between immune cells and their activation, migration, proliferative expansion, and differentiation is orchestrated by a broad range of cytokines including interleukins and chemokines. Some cells (like ILCs) are directly activated by interleukins, whereas others (like T cells) require antigenic stimulation to upregulate surface receptors, which then can mediate functional responses. For obvious reasons, "differentiation" is a continuous ongoing process, which governs the immune system from the early steps of embryonic development all the way to the regulation of immune responses in the mature immune system. All steps of immune cell development and differentiation are subject to epigenetic regulation.

Development of immune cells

All immune cells develop from a common hematopoietic stem cell (HSC), which gives rise to lineage-specific progenitor cells, i.e., common lymphoid progenitor (CLP) and common myeloid progenitor (CMP) cells. While these cells still maintain the potential for (unlimited) self-renewal, their capacity for differentiation is restricted to distinct cell lineages. CLP are the precursors of B cells, T cells, ILCs, and NK cells, whereas CMP give rise to granulocytes and monocytes/macrophages on one side and to immature dendritic cells (DCs) on the other side (Fig. 1). A detailed overview of immune cell development is available in standard text books [1].

T cells develop from precursor cells entering the thymus where they undergo sequential maturation steps characterized by lack of CD4 and CD8 expression ("double negative," DN) but differential expression of CD44 and CD25. Rearrangement of the TCR β variable (V), diversity (D), and joining (J) gene segments occurs at the $\mathrm{CD44^{low}CD25^{high}}$ DN3 stage, where also the branching of the second population of T cells expressing a γδ TCR instead of the conventional αβ TCR takes place [2]. Following the subsequent rearrangement of TCR Vα and Jα genes, randomly generated TCR αβ heterodimers are expressed on the cell surface, which then undergo positive and negative selection based on affinity of their TCRs for self-major histocompatibility complex (MHC) class I or class II molecules in the context of self-peptides available in the thymus. The autoimmune regulator (AIRE) plays a critical role in this process by controlling the expression of peripheral tissue antigens in medullary thymic epithelial cells [3]. Overall, thymic selection of T cells is associated with massive cell death. Thymocytes expressing TCRs with no or too low affinity for self-MHC die by neglect, whereas thymocytes expressing TCRs with high affinity for self-MHC (which might cause fatal autoimmunity if exported to the periphery) die by apoptosis. This process, termed central tolerance, ensures that the vast majority of T cells leaving the thymus express TCRs of intermediate affinity for self-MHC, which will allow them

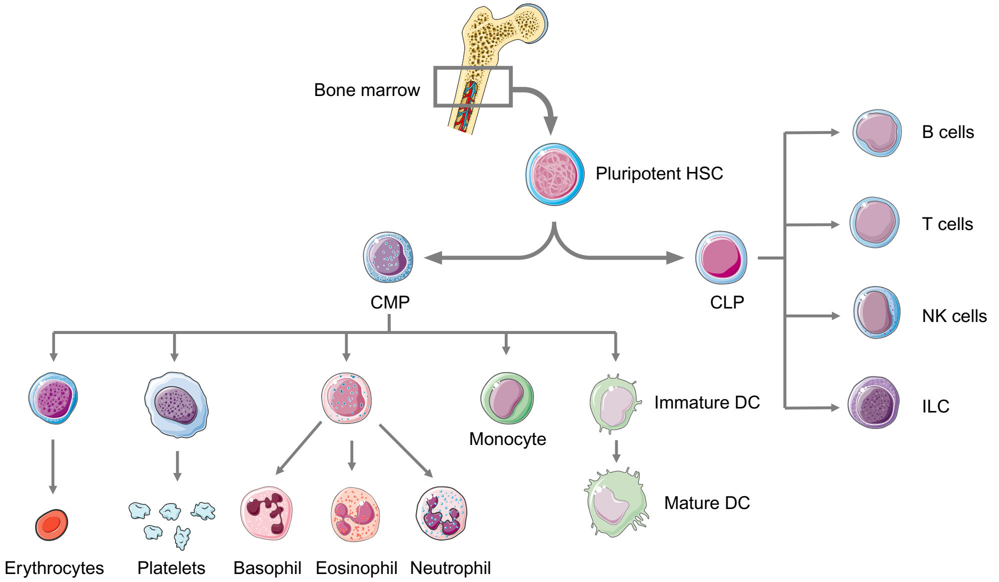


Fig. 1 A cartoon of immune cell development. Immune cells develop from hematopoietic stem cells (HSC), which give rise to lineage-restricted progenitor cells, i.e., common lymphoid progenitors (CLP) and common myeloid progenitors (CMP) cells. CLP cells are progenitors of adaptive immune cells expressing clonally variable antigen receptors (B cells/BCR, T cells/TCR) and innate lymphocytes lacking antigen receptors (NK cells, innate lymphoid cells ILC). Common myeloid progenitors give rise to both granulocytes/monocytes/macrophages and immature dendritic cells (DC), which can mature into potent antigen-presenting cells. Once blood-borne monocytes migrate into tissue they differentiate into macrophages. Tissue-specific macrophages include Langerhans cells in the skin, Kupffer cells in the liver, and microglia in the brain.

to recognize processed peptides in the context of MHC class II (CD4 T cells) or MHC class I (CD8 T cells) molecules. Central tolerance, however, does not work perfectly, and peripheral mechanisms are instrumental for the maintenance of self-tolerance. Among those, T regulatory cells (Treg) have a crucial function. Treg are characterized as $CD4^{+}CD25^{high}CD127^{low}$ and express the specific transcription factor FoxP3. FoxP3-deficient mice develop lymphadenopathy and succumb to autoimmunity, and *FOXP3* mutations in humans giving rise to immunodysregulation, polyendocrinopathy, enteropathy, X-linked (IPEX) syndrome are similarly associated with autoimmune phenomena [4]. Demethylation of Treg-specific demethylated regions (TSDR) in the *FOXP3* gene is required for the suppressive activity of Treg [5].

B-cell producing antibodies of the immunoglobulin isotypes IgM, IgG, IgD, IgE, and IgA develop in the fetal liver and bone marrow. Similar to T-cell receptors, antibody diversity is generated by V(D) J recombination of heavy (H) and light (L) chain genes during B-cell development. Following the expression of B-cell specifying genes like E2A

(*E*-box binding protein) and EBF (early B-cell factor), VDJ recombination is initiated at the Ig heavy chain locus. The B-cell receptor (BCR) consists of membrane-bound Ig, which mediates antigen recognition, and the associated Ig-α/Ig-β heterodimer, which comprises the BCR signaling subunit. Signaling via the antigen receptors of adaptive immune cells (BCR and TCR, respectively) requires immunoreceptor tyrosine-based activation motifs (ITAM), which are present not only in the BCR-associated Ig-α/Ig-β molecules but also in the TCR-associated CD3 subunits. B-cell activation and differentiation take place in the germinal centers in lymph nodes, where naïve B cells upon encounter of specific antigen undergo proliferation and somatic hypermutation in V gene regions leading to affinity maturation of antibodies during an ongoing immune response. An important feature of B-cell biology is the Ig class switch taking place during an ongoing immune response. Surface IgM-positive B cells start to secrete IgM but switch to IgG and, depending on specific signals, to IgA or IgE. Class switching is promoted by specific cytokines such as IL-4 for IgE or transforming growth factor-β (TGF-β) for IgA. While plasma cells secrete large amounts of antibodies, some cells survive in niches as long-lived plasma cells and memory B cells, ready to be reactivated upon later encounter of the same antigen [6]. All steps of B-cell activation and differentiation are influenced by epigenetic regulation [7].

Natural killer (NK) cells have been discovered over 40 years ago as lymphoid cells lacking typical markers of B cells and T cells endowed with the capacity to recognize and kill tumor cells and virus-infected cells without prior sensitization. Like T and B cells, NK cells originate from CLP cells. Today, NK cells are classified as potent cytotoxic and cytokine-producing cells whose activity is regulated through the concomitant activity of a plethora of both activating and inhibitory receptors. Some activating receptors like the natural killer group 2 member D (NKG2D) receptors sense stress-inducible MHC class I-related molecules (in humans MHC class I-related chain A/B, MICA/B; and UL-16-binding proteins, ULBP1–6), while others like the natural cytotoxic receptors NKp40 recognize viral hemagglutinins. Killer inhibitory receptors (KIR) on the other hand recognize MHC class I molecules. NK cells can kill target cells via the direct release of lytic granules or by inducing death receptor-mediated apoptosis [8]. Overall, the NK cell system is tuned in such a way that NK cells are not activated by healthy cells expressing "normal" levels of MHC class I molecules, but get activated when recognizing cells with reduced MHC class I expression as it occurs upon viral infection of cellular transformation ("missing self" hypothesis put forward by Klas Kärre) [9].

Another class of lymphoid cells lacking antigen-specific receptors but originating also from CLP cells is the innate lymphoid cells (ILCs). ILCs are mostly tissue-resident and can be subgrouped according to the expression of transcription factors and correlated effector functions. Group 1 ILCs (ILC1) are activated by epithelia-derived IL-15, express the transcription factor Tbet, and produce interferon-γ (IFN-γ), which augments defense against intracellular bacteria and viruses. ILC2 are activated by IL-25 and IL-33, express GATA3, and produce type 2 cytokines like IL-5 and IL-13 important for defense against helminths and for regulating allergy and asthma. ILC3 are activated by IL-7, express

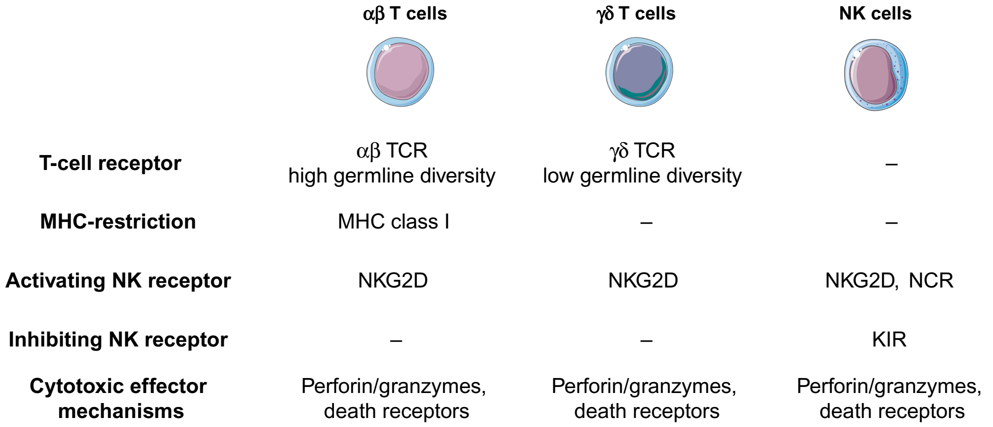
RORγt, and secrete IL-22 and IL-17 involved in defense against extracellular bacteria and fungi, and also in autoimmunity [10]. NK cells are closely related to ILC1 in terms of transcription factor expression and IFN-γ production, but additionally exert cytotoxic activity that is absent in ILC1. Another lymphoid population with close links to ILC3 (also depending on RORγt) are the lymphoid tissue-inducer (LTi) cells crucial for the formation of secondary lymph nodes and Peyer's patches during embryonic development. The functional dichotomy of ILCs, i.e., the correlation between the expression of a specific transcription factor with the range of produced effector cytokines, is reminiscent of the similar functional subgroups of conventional CD4 and CD8 T cells. However, ILCs are profoundly involved in tissue homeostasis, particularly in the intestine, lung, and adipose tissue [11].

Dynamics of the immune response

The immune response to infection involves a continuum of innate and adaptive mechanisms [1]. Physiologically, the barriers to the outside world (like skin, mucosa in the intestine and lung) are protected by naturally occurring or inducible antimicrobial peptides such as defensins and others [12]. Once microbes cross the epithelial barrier, tissue-resident macrophages and innate dendritic cells (DC) are alerted by sensing the danger through pattern recognition receptors (PRRs). Some PRRs are located on the cell surface, e.g., C-type lectin receptors, scavenger receptors, and Toll-like receptors (TLR), which recognize cell wall constituents of Gram-positive (TLR1/2/6) and Gram-negative (TLR4) bacteria, or specific proteins of flagellated bacteria (TLR5). Other PRRs are located intracellularly, such as NOD-like receptors (NLRs), the cytosolic RNA receptor RIG-I, the cytosolic dinucleotide receptor cGAS/STING, endosomal TLR-sensing single- (TLR7/8) or double-stranded (TLR3) viral RNA, and hypomethylated CpG motifs of bacterial DNA (TLR9) [13]. Upon uptake of microbial antigen and stimulation of pattern recognition receptors, the ensuing signal transduction (frequently culminating in the activation of the pro-inflammatory transcription factor NF-κB) induces maturation of DCs and their migration to the next available lymph node where encounter with B and T cells searching for "correct" antigens via their BCR and TCR takes place. Upon maturation, DCs upregulate MHC class I and class II molecules as well as ligands for co-stimulatory receptors on T cells such as CD80 and CD86, making mature DCs the most potent antigen-presenting cells. DCs can take up antigen by phagocytosis, pinocytosis, and receptor-mediated endocytosis. Upon internalization, antigen proteolysis takes place in lysosomes, and peptides are loaded onto MHC class II molecules for presentation to CD4 T cells. Peptides for loading onto MHC class I molecules are generated from the degradation of cellular or viral antigens in the proteasome. They are subsequently transported into the endoplasmatic reticulum where uploading takes place for presentation to CD8 T cells [14]. Importantly, DCs can also channel peptides from endocytosed antigens into the MHC class I presentation pathway (a process termed cross-presentation) [15]. Apart from DCs, activated human γδ T cells are the only non-professional antigen-presenting cells endowed with the capacity for cross-presentation [16]. While mature DCs are the most potent antigen-presenting cells for the activation of CD4 and CD8 T cells, immature DCs lacking the expression of co-stimulatory molecules are in fact important regulators of antigen-specific tolerance [17].

CD4 T cells recognizing via their clonally distributed TCR the "correct" antigenic peptide presented on DCs by MHC class II get activated, i.e., signal transduction is initiated by tyrosine phosphorylation of CD3ζ chain ITAMs and recruitment and phosphorylation of adaptor molecules like lck (signal 1). However, in the absence of signal 2, the co-stimulation provided by the binding of CD28 on T cells to corresponding ligands CD80 or CD86 on antigen-presenting cells, T cells do not get fully activated but rather remain anergic. In addition to CD28, other receptor/ligand interactions can also provide signal 2 [18]. Signal 3 of T-cell activation is delivered by cytokines like IL-2, which trigger cell division of activated T cells. Importantly, during activation, T cells also upregulate inhibitory receptors like CTLA4 and PD1, which serve to terminate cellular immune responses upon completion of the task. Interfering with these inhibitory signals by antibodies against the inhibitory receptors (or the ligand in case of PDL-1) has developed into a clinically efficient strategy to unleash antitumor T-cell responses in certain types of cancer [19]. Activated CD4 T cells can exert a broad range of effector functions imprinted by the micromilieu encountered during activation. CD4 T cells can thus differentiate into Th1, Th2, Th17, Th9 subsets characterized by the expression of lineage-associated transcription factors (like ILCs, see above) Tbet, GATA3, RoRγt, and PU.1, respectively, and corresponding key cytokines IFN-γ, IL-4/IL-5/IL-13, IL-17/IL-22, and IL-9, respectively. An oversimplified categorization implies that Th1 cells are important for the activation of macrophages and for the defense of intracellular bacteria, Th2 cells for the defense of helminths, allergic reactions, and B-cell differentiation, Th17 cells for regulating autoimmunity and inflammation, and Th9 cells for defense against extracellular bacteria and tumor immunity. T-cell differentiation in response to environmental signals is heavily regulated by epigenetic processes [20, 21].

CD8 T cells expressing αβ TCR play a major role in the elimination of virally infected cells. They recognize in an MHC-restricted fashion peptide derived from viral proteins bound to the peptide pocket of MHC class I molecules. Similarly, cells expressing altered self-MHC class I molecules or tumor-associated antigens can be recognized by CD8 T cells. CD8 T cells use two strategies to kill target cells: (i) the release of cytolytic granules with pore-forming perforin and serine proteases like granzyme B and (ii) by inducing apoptosis in death receptor (e.g., Fas/CD95 or TRAIL-receptor)-expressing target cells through induced expression of corresponding ligands (Fas-L or TRAIL). As discussed earlier, CD8 αβ T cells possess an almost unlimited TCR repertoire and are characterized at the clonal level by exquisite specificity for defined peptides in the context of self-MHC class I molecules. As mentioned earlier, there is a small subset of T cells expressing a different TCR composed of γ and δ chains. There is overwhelming evidence that γδ T cells are important players in the local immune surveillance [22]. Moreover, γδ T cells contribute broadly to antiinfective and antitumor immunity [23]. Human γδ T cells do not recognize peptides in a MHC-dependent manner but rather recognize phosphorylated intermediates of the mevalonate pathway of cholesterol synthesis. Such pyrophosphates activate all human γδ T cells expressing the Vγ9Vδ2 TCR, which in most healthy adult donors account for the majority (50% to >95%) of the 2%–5% γδ T cells present in peripheral blood. In recent years, it was discovered that members of the butyrophilin transmembrane molecule family, specifically the BTN3A1 isoform, play an indispensable role in this process [24]. Importantly, many tumor types have a dysregulated mevalonate pathway and produce increased amounts of pyrophosphates which then activate γδ T cells. Since the recognition of pyrophosphates is not MHC restricted, activated human γδ T cells can recognize and kill a variety of human tumor cells that forms the basis for novel therapeutic strategies [25]. Again, the activation of γδ T cells is modulated by epigenetic regulation [26, 27]. Taken together, at least three classes of cytotoxic lymphocytes can contribute to the antitumor immune response, i.e., MHC class I–restricted CD8 αβ T cells, MHC nonrestricted γδ T cells, and NK cells. A comparative overview of their main features is presented in Fig. 2.



	αβ T cells	γδ T cells	NK cells
T-cell receptor	αβ TCR high germline diversity	γδ TCR low germline diversity	–
MHC-restriction	MHC class I	–	–
Activating NK receptor	NKG2D	NKG2D	NKG2D, NCR
Inhibiting NK receptor	–	–	KIR
Cytotoxic effector mechanisms	Perforin/granzymes, death receptors	Perforin/granzymes, death receptors	Perforin/granzymes, death receptors

Fig. 2 Characteristics of three classes of cytotoxic lymphocytes. CD8 αβ T cells and γδ T cells are T lymphocytes which express a CD3-associated T-cell receptor (TCR). The TCR repertoire of CD8 αβ T cells is extremely high, whereas the TCR repertoire of γδ T cells is limited. CD8 αβ T cells recognize peptides in an MHC class I-restricted manner, whereas γδ T cells recognize nonpeptide molecules (e.g., pyrophosphates in the case of human Vγ9Vδ2 T cells) in an MHC nonrestricted manner. NK cells lack CD3 and TCR but express a broad range of inhibitory (e.g., killer inhibitory receptors, KIR) and activating (e.g., natural killer group 2 member D, NKG2D; Natural Cytotoxic Receptors, NCR; e.g., NKp46) receptors. Some of the NK receptors like NKG2D can be also expressed on cytotoxic CD8 αβ and γδ T cells. All cytotoxic lymphocytes can use secretory pathway (release of perforin and granzymes) and/or death-receptor pathway (e.g., Fas/Fas-ligand, TRAIL-R/TRAIL) to kill target cells.

Immunological memory

Long-lived memory is a hallmark of the immune system, and the induction of memory forms the basis of successful vaccination against infectious diseases. In the B-cell compartment, both memory B cells and long-lived memory plasma cells contribute to the establishment of memory. Long-lived plasma cells survive in specific niches in the bone marrow and spleen, and are not depleted during B-cell depleting antibody therapies as applied in severe autoimmune diseases [28, 29]. When compared to a primary (mostly IgM) response, memory cells enable the rapid production of increased amounts of protective (mostly IgG) antibodies upon antigenic rechallenge [6]. Within the CD4 T-cell compartment, several surface markers have been identified over the years which correlate with the level of previous antigen experience ("priming") and the extent of memory response. Based on the expression of CD27, CD45RA, and the chemokine receptor CCR7, subsets of human CD4 T cells are classified as naïve ($CD27^+CD45RA^+CCR7^+$), central memory T_{CM} ($CD27^+CD45RA^-CCR7^+$), or effector memory T_{EM} ($CD27^-CD45RA^-CCR7^-$). While such a classification helps to understand the dynamic phenotypic alterations of CD4 T cells during activation and differentiation and their relevance for memory development, it is obvious that the cellular differentiation at the clonal level is much more complicated and involves the microenvironment-dependent regulation of multiple cell surface antigens [30]. More recently, it has been recognized that there are sizeable numbers of tissue-resident memory T cells (T_{RM}) of both CD8 and CD4 subsets which reside in the lung, skin, and other mucosal tissues without recirculating. They are recruited from circulation and secondary lymphoid organs following infection, and typically express high levels of CD69 [31].

Until recently, it was thought that memory is an exquisite feature of the adaptive immune system, i.e., restricted to B and T cells. It is quite clear, however, that innate immune cells can also "remember" previous exposure to microbial stimuli (Fig. 3). NK cells can mount a robust recall response during viral infection which is induced by cytokines and NK receptors [32, 33]. Importantly, upregulation of the activating receptor NKG2C has been found to confer increased specificity for human cytomegalovirus (HCMV) because NKG2C can recognize HMCV-specific peptides—a surprising example of antigen specificity of antigen-nonspecific innate NK cells [34]. Other innate immune cells that are amenable to memory responses are monocytes and macrophages. Like NK cells, monocytes/macrophages do not express somatically rearranged antigen receptors, but do express a variety of PRRs sensing molecular patterns like cell wall constituents (e.g., LPS of Gram-negative bacteria, peptidoglycans of Gram-positive bacteria). Monocytes/macrophages exposed to β-glucan, *Candida* infection, or Bacillus Calmette Guérin (BCG) vaccination are more strongly activated and secrete increased amounts of pro-inflammatory cytokines upon rechallenge with other microbes, a process which has been termed trained immunity [35]. Trained immunity is a primitive form of

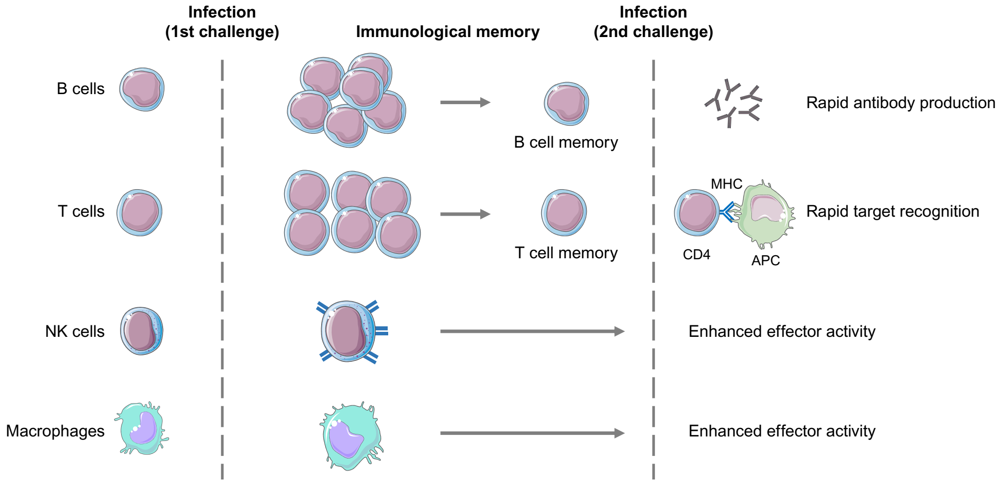

Infection
(1st challenge)
Immunological memory
Infection
(2nd challenge)
B cells
B cell memory
Rapid antibody production
T cells
T cell memory
MHC
CD4
APC
Rapid target recognition
NK cells
Enhanced effector activity
Macrophages
Enhanced effector activity

Fig. 3 Memory in the adaptive and innate immune system. Antigen-specific B cells and T cells undergo clonal expansion upon antigen encounter. During an ongoing immune response, long-lived memory cells develop which enable rapid recall responses upon repeated challenge with the same antigen. B-cell memory comprises memory B cells and long-lived memory plasma cells which reside in specific niches in the bone marrow. Central memory (T_{CM}) and effector memory (T_{EM}) CD4 T cells circulate between blood and secondary lymphoid organs. Tissue-resident memory T cells (T_{RM}) reside in tissues and serve as first line of defense at barrier sites. Memory is not restricted to the adaptive arm of the immune system. NK cells can acquire memory leading to enhanced activity upon rechallenge with (virus)-infected cells. Upon infection, innate immune memory can develop in monocytes and macrophages through the upregulation of pattern recognition receptors, thereby leading to increased nonspecific responses during subsequent infections.

adaptation of the host to infection, and it is increasingly recognized that development of immunological memory is a continuous process, which is based on intertwined mechanisms of innate and adaptive immunity [36]. The molecular mechanisms behind adaptation to environmental stimuli and the development of immunological memory are certainly complex and context-dependent. The intensity and quality of receptor signaling, metabolic reprogramming, chromatin/histone modifications, and epigenetic modification at the level of DNA methylation are just a few important parameters [37]. Overall, it is obvious that epigenetic regulation is very important for the orchestration of physiological immune responses and maintenance of immunological memory.

Basic overview of epigenetic regulation

Addressing the view on classical controversies in embryology during the early 20th century, Conrad Waddington wrote in his book, "An introduction to Modern Genetics," "… but equally it is clear the interaction of these constituents (of the fertilized eggs) gives rise to new type of tissue and organ which were not present originally, and in so far development must be considered as 'epigenetics'" [38, p. 156]. Thus, he introduced the term "epigenetics" for such nongenetic mechanisms as "the branch of biology which studies the causal interactions between genes and their products, which bring the phenotype into being" [38].

Throughout the last decades scientists have come up with different views on and various definitions of epigenetics. The current definition of epigenetics has evolved as, "the study of changes in gene function that are mitotically and/or meiotically heritable and that do not entail a change in DNA sequence" [39]. Nonetheless, epigenetics, in a broader sense, is a bridge between genotype and phenotype. For instance, multicellular organisms exhibit identical genotypes, but the functional diversity of processes specifies a context-depending unique cell type. This tremendous capacity of (cellular differentiation by) epigenetic mechanisms controls heritable changes in gene expression or cellular phenotype and is described as "epigenetic landscape" [40]. The major breakthrough came during the period from 1869 to 1951, when Miescher, Flemming, Kossel, Heitz, Muller, and McClintock laid the foundation of modern era of epigenetics, by providing early hints for non-Mendelian inheritance. Specifically, the pioneering study by the German scientist Emil Heitz formed the cytological basis for the distinction between euchromatin (genetically active) and heterochromatin (genetically inactive) [41]. So, he further classified heterochromatin as constitutive heterochromatin, where both maternal and paternal chromosomes respond in similar way during development, and as facultative heterochromatin, where both homologous chromosomes behave differently (one becomes heterochromatic and the other remains euchromatic) [42].

The epigenetic mechanisms can be studied through, but not limited, two different perspectives, i.e., structural and functional epigenomics. In the following sections, we have discussed both perspectives to understand the functional regulation of gene expression.

Genome architecture at primary structure scale: Cis-regulation through DNA elements

The eukaryotic gene expression and regulation have been targeted and manipulated to study the function of the respective gene at the protein level. However, the genomic locations play an important role in such regulatory processes. In a cis-regulation of gene expression, promoter regions play important roles. A promoter region is composed of core promoter and proximal promoter elements and spans nearly 1 kb pairs. Traditionally, the promoter region is considered as the key region controlling gene expression as it facilitates the binding of the preinitiation complex and the transcriptional machinery. However, recent reports on the involvement of distal regulatory elements show that these elements play more active roles in a spatial or temporal manner, independent of distance and orientation, which leads to cis- and/or trans-regulation of gene expression. Distal regulatory elements are enhancers, silencers, insulators, and locus control regions. Enhancer regions have clusters of transcription factor binding sites and functionally work similar to proximal promoter elements, which enhance transcription. Silencers are sequence-specific elements, which allow binding of transcriptional repressors, resulting in repressed transcription. Insulators or boundary elements are ~0.5–3 kb pairs in length and function in position-dependent manner to block genes from being affected by transcriptional activity of neighboring genes and thus partition the genome. Insulators possess two main properties: (i) enhancer-binding activity and (ii) heterochromatin-barrier activity. Locus control regions (LCRs) are the multiple regulatory elements, including cis-acting elements like enhancers, silencers, insulators, and nuclear-matrix or chromosome scaffold-attachment regions (MARs or SARs). LCRs function together to regulate the entire locus or gene cluster [43]. Thus, the promoters and distal regulatory elements control expression of an mRNA transcript, encoding exon-spanning sequences of genomic region. This cis-regulation of gene expression has been studied for a long time to understand the role of a protein in immune cell function (Fig. 4A).

The application of single-nucleotide resolution studies (e.g., whole-genome bisulfite sequencing—WGBS or its alternative methods like reduced representation bisulfite sequencing—RRBS) along with sequencing methods such as ChIP-seq, ATAC-seq is necessary to understand the primary structure of genomic architecture [44]. One of the important and most studied components of this single-nucleotide level gene regulation is DNA methylation. DNA methylation is primarily an indicator of transcriptional repression, which depends on the presence of active methylation site, CpG islands and their interaction with the transcriptional machinery. DNA demethylation also referred to as DNA hydroxymethylation, is essentially regulated through the ten-eleven translocation (TET) enzymes and also plays an important role in the expression and regulation of transcription factors responsible for cellular identity [45].

Genome architecture at secondary structure scale: Trans-regulation through chromatin packaging and nucleosome positioning

A cell is the most basic building unit of life and holds all the genetic information in a total of 6 million base pairs of DNA. A 2-nm-wide DNA double helical structure is called chromatin. Typically, chromatin is made up of octamer histone proteins (H2A, H2B, H3, and H4) wrapped around 146 bp of DNA (or 1.7 turns of DNA) and linker DNA of about 20–60 bp forming the nucleosome [46]. This structure is often referred to as "string of beads." A nucleosome with an H1 histone forms a "chromatosome," an 11-nm structure. Next, these chromatosomes fold up together to produce a 30-nm fiber that forms loops with an average length of 300 nm. These 300-nm fibers are further coiled to produce "the chromatid of a chromosome." Coiling of "string of beads" and fibers is important to compact 2-m-long DNA into the microscopic space of the eukaryotic nucleus. Positively charged small histone proteins bind very tightly to negatively charged DNA (because of the phosphate group within the phosphate-sugar backbone) by providing the energy mainly in the form of electrostatic interactions [47] (Fig. 4B).

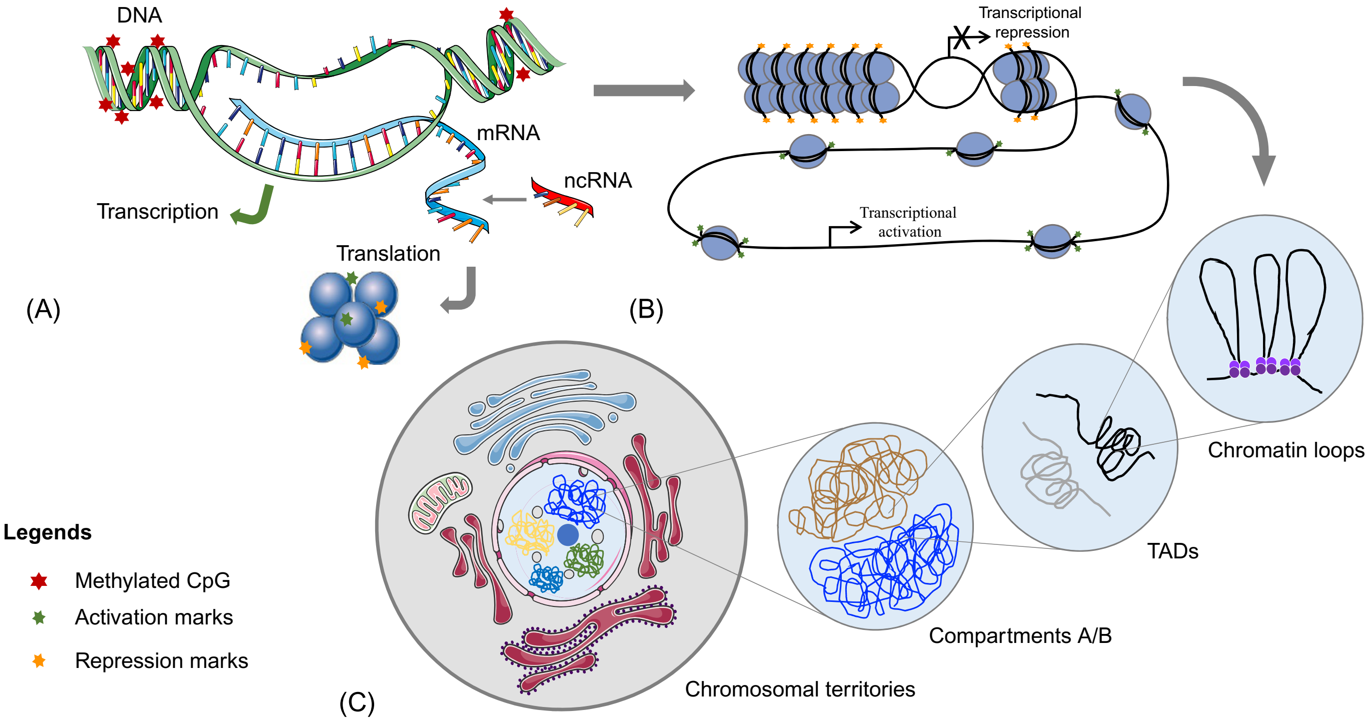

DNA
mRNA
ncRNA
Transcription
Translation
(A)
Transcriptional repression
Transcriptional activation
(B)
Chromatin loops
TADs
Compartments A/B
Chromosomal territories
(C)
Legends
Methylated CpG
Activation marks
Repression marks

Fig. 4 Genome organization and associated multilayers of epigenetic mechanisms in the cell. A simplified overview of genome organization from a single base-pair resolution to 10 Mb long compartments and territories. (A) In the cis-regulation, DNA is transcribed to RNA due to the interaction of transcriptional machinery at the regulatory regions. This process is epigenetically regulated by posttranscriptional and posttranslational modifications including DNA methylation (5 mC) and noncoding RNA (ncRNA). In addition to this single-base pair resolution process, the nucleosome positioning and chromatin accessibility (B) play an important role at the nucleosomal scale. (C) At the next level of supranucleosomal scale in hierarchical genome organization, the formation of chromatin loops, topologically associated domains (TADs), and chromatin compartments A/B controls important molecular mechanisms. In the highest order of hierarchy, i.e., at the nuclear scale, the chromosomal territories are formed within nuclear space. Abbreviations: *mRNA*, messenger RNA; *ncRNA*, noncoding RNA; *TADs*, topologically associated domains.

Histone modification, chromatin accessibility, and nucleosome remodeling are primary regulators of gene expression [48], which can be studied using relatively new techniques such as ChIP-seq, ATAC-seq, DNase-seq, MNase-seq, FAIRE-seq, and so on.

Genome architecture at tertiary structure scale: Long-range chromatic interactions and gene regulation within chromatin territory and compartments

In recent years, ground-breaking reports have been published to unravel the mechanism of genome folding and transcriptional regulation in the interphase nuclear space. The tertiary structure of genomic architecture comprises dynamic cellular and molecular processes. The genetic material is localized in the nucleus, surrounded by the porous nuclear envelope, actively involving transcriptional processes such as transport of RNA [49]. The nuclear envelope with two membranes (especially inner side) plays an important role in nuclear genomic organization. In the nucleoplasm, nucleolus, Kajal bodies, Splicing bodies, and promyelocytic leukemic bodies are presumably self-organizing structures. Consequently, the chromatin architecture is organized at multiple levels, defining "chromosomal territories," within the nucleus and their spatiotemporal positioning and having its correlated gene content and activity. These multiple levels of organization arguably include transcriptionally inactive chromatin near nuclear chromocenters, pericentromere-associated domains, near nucleolus is so-called nucleolus-associated domains, nuclear lamina-associated domains (LADs), chromatin compartments (A & B), topologically associated domains (TADs), and chromatin loops [50]. Thus, the human genome folds and creates thousands of intervals with enhanced contact frequencies exhibiting contact, also referred to as "contact domains" [51]. Though the knowledge of the higher order chromatin organization is still limited, all the above terms are being proposed in the form of mainly two models: hierarchical model and alternative model of genome organization [52]. The hierarchical model supports that differently sized features play a role in each other's formation. The model explains that at the order of about 1–10 Mb, each chromosome can be divided into two "compartments," i.e., A and B with an exclusive chromatin interaction pattern. The compartment A is transcriptionally active and rich in gene content and activity while B is transcriptionally inactive and poor in gene content and activity. The next level in the hierarchy is the smaller chromatin structure of self-interacting TADs that have insulated genomic regions of about 0.5–1 Mb in size. These interacting regions up to 1 Mb are also referred to as chromatin "loops." These chromatin loops can be presumably divided as structural loops and regulatory loops. Sub-TADs are characterized by higher interaction frequencies while structural loops have strong interactions between CTCF sites at the borders of TADs. The regulatory loops form contacts between gene promoters and regulatory elements such as enhancers and super-enhancers, additionally bound by structural factors like YY1 [50]. Additionally, YY1 and CTCF tether gene loci to nuclear lamina or nucleolus. Thus, additional factors such as RNA polymerase II (hereafter referred to as "pol II"), mediators, and bromodomain-containing protein 4 are associated with active chromatin (euchromatin) and factors like polycomb proteins, heterochromatin proteins, pol II depleted regions are associated with inactive chromatin (heterochromatin) [53, 54]. Of note, the loop extrusion model for chromatin loop formation has been widely studied. Formation of the loop is a cohesion-dependent process, mediated by cohesin and CTCF. Cohesin, as a tripartite ring of Scc1-Smc3-Smc1, primarily holds a loop together with CTCF in a specific orientation (Fig. 4C). Thus, the loop-formation process per se is a cohesin-dependent, but not affected by CTCF depletion [55]. Chromatin states are additionally segregated by the presence of different molecules (e.g., multivalent proteins) creating molecular class-specific interactions among the compartments. Such creation of phase-separated multivalent, multimolecular assemblies and their transcriptional regulation is referred to as the "phase separation model for transcriptional control" [52].

Posttranscriptional regulation of RNA expression

Regulation of gene/protein expression occurs at three different stages: first at the level of transcriptional regulation during the generation of mRNA from DNA, second at the level of translational regulation during the generation of gene product from mRNA, and third by posttranscriptional or posttranslational regulation through various modifications. RNA synthesis or transcription is a highly coordinated process that includes four different steps, i.e., recognition, initiation, elongation, and termination. This transcriptional program is often regulated through the interplay between chromatin remodeling complexes [e.g., bromodomain (Brd) proteins including Brd and ExtraTerminal domain (BET) proteins] and other epigenetic elements, and their binding specificity to DNA (coding and noncoding) sequences [56]. Traditionally, the promoter regions are considered to play an important role in the binding of transcription factors and regulatory proteins. Enhancers located in noncoding DNA gain more importance in the very same steps of recognition and initiation [57]. As epigenetic modifications play a central role in the control of gene expression, histone modifications can be used to define the regulatory regions and forms. H3K4me3 (representing active state) are positively correlated with gene expression, while negatively correlated with H3K27me3 (representing repressive state) in the cell. Interestingly, based on these correlations, gene expression together with histone methylations are further divided into four distinct forms: active, repressive, poised, and bivalent. When considering only gene expression, three distinct patterns are identified: active, repressed, and poised [58]. Further adding to the complexity, it was found that the process of transcription elongation is highly regulated by pol II pausing mechanisms, which coordinate other potential epigenetic mechanisms like nucleosome positioning and noncoding RNA [59, 60]. Pol II pausing may govern a primary role in the co-expression and co-regulation of genes. Thus, considering the abundance of pol II at promoter regions of specific gene(s) and expression/regulation of other active gene(s), these specific genes are also called "immediate early genes," "late response genes" (further divided as delayed-primary response genes and secondary response genes), and "universal amplifier genes." Some examples of these genes are as follow: immediate early genes—c-Jun, c-Fos, delayed-primary response genes—VCL, PLOD2, secondary response genes—MMP3, MMP13, and universal amplifier genes—c-Myc [61, 62].

Fully functional mRNA transcripts are produced by passing through above said essential steps and additional RNA processing steps such as splicing, capping, and polyA tail addition. This mRNA can be used to construct proteins, hence is referred to as protein-coding RNA. However, this coding region represents merely 3% of the total genome. Importantly, the big portion (~62%) is transcribed into regulatory noncoding RNA (ncRNA) [63]. Based on their dimension, regulatory ncRNA are classified as "small" ncRNA (<200 nucleotides) and "long" ncRNA (>200 to >1000 nucleotide). These families of regulatory ncRNA are summarized in Table 1 [64, 65]. The role of ncRNA has been described in a context-dependent manner during events of chromatin conformation, histone modification, posttranscriptional processing, localization, and degradation [66, 67]. More than 100 types of posttranscriptional RNA modifications have been identified, of which N6-methyladenosine (m6A) RNA methylation is one

Table 1 Main classes of regulatory noncoding RNA.

Abbreviation	ncRNA	Length (nt)	No. of known transcripts
Short ncRNA			
miRNA	MicroRNA	21–23	1756
snoRNA	Small nucleolar RNA	60–300	1521
snRNA	Small nuclear RNA	150	1944
piRNA	Piwi-interacting RNA	25–33	–
PASR	Promoter-associated short RNA	22–200	–
TASR	Termini-associated short RNA	22–200	–
siRNA	Short interfering RNA	21–23	–
tiRNA	Transcription initiation RNA	15–30	–
tRNA	Precursors to short transfer RNA	73–93	497
Long ncRNA			
NAT	Natural antisense transcripts	>200	5446
PALR	Promoter-associated long RNA	200–1000	–
PROMT	Promoter upstream transcript	200–600	–
T-UCR	Transcribed ultraconserved regions	>200	–
Intronic RNA	Intronic RNA	>200	–
eRNA	Enhancer-derived RNA	>200	>2000
LincRNA	Long intervening/intergenic RNA	>200	6742
uaRNA	3′UTR-derived RNA	<1000	12
circRNA	Circular RNA	100 to >4000	–

of the most prevalent modifications of mRNA, also mediated by miRNA [68]. In addition, the alternative splicing and expression quantitative trait loci (eQTL) play pivotal roles in the context-dependent regulation of mRNA expression [69].

Posttranslational modifications: Chromatin remodeling complexes and histones

The organization of chromatin and nucleosome is a structurally dynamic process. This dynamic process is not only important for the packaging of DNA, but also for the regulation of DNA accessibility for transcription, recombination, DNA repair, and replication, which ultimately leads to the cumulative effect of gene expression during health and/or disease progression. These complex processes are mainly regulated by energy-dependent chromatin remodelers, histone chaperons, and histone-modifying enzymes. Based on structural and functional domain outside the enzymatic core, the so-called SWI/SNF, ISWI, CHD, and inositol-requiring 80 (INO80) protein families are the four groups of chromatin remodelers [70]. Chromatin remodelers use the energy of ATP hydrolysis to drive a DNA translocase. This allows nucleosome repositioning by sliding, eviction, assembly, spacing, or histone replacement in order to maintain higher order chromatin organization and to modulate its accessibility. These accessible regions are possibly available for histone chaperones to carry out nucleosome assembly or for histone exchange of variants to regulate (post) transcriptional processes. Histone chaperones are histone-interacting proteins that play a crucial role in transcriptional regulation, histone storage, transport, nucleosome stability, assembly, and disassembly. They are classified based on the type of substrate they bind. Histone chaperones are shown to bind to H3-H4, H2A-H2B oligomers or both hetero-oligomers. Depending on their location and/or function, they can bind to specific canonical or variant histones alone. Serving as important linkers with chromatin remodelers, they cooperate to act as "histone sinks," also known as "histone acceptors" [71]. But, the basic question is, "what are histones?"

Basically, histones are highly alkaline proteins found in eukaryotic cells, which serve as main components of chromatin. Histone proteins are distinguished into two distinct ways: one being canonical (or core) histones (H2A, H2B, H3, and H4) and the other their histone variants. The core center of the nucleosome is an H3/H4 tetramer, due to strong four helix bundle interaction between the two H3 proteins. These H3/H4 tetramer interact with H2A/H2B and thus provides docking site for entry-exit of DNA via H2A C terminus domain. H2A and H2B turnover are higher than those of H3/H4 due to weaker intranucleosomal contacts, sensitivity to transient unwrapping, and DNA exposure [72]. Gene expression of histone genes is very important and highly regulated, but different for canonical and variant histones. Canonical histone synthesis and deposition is dependent on DNA synthesis. These genes are clustered and transcribed during the S phase of cell cycle. A unique feature is that mRNA lacks introns and poly A tails, which are replaced with special regulatory stem-loop structure to stimulate translation. Fully synthesized canonical histones are deposited into nucleosome behind the replication fork, at the site of DNA repair. Histone variant gene expression is DNA-independent and employs pol II transcripts. As per evolutionary analysis, all H3 and H2A can be considered as histone variants too [73]. Histone variants possess distinct amino acid sequences, which influence the physical properties of nucleosome dynamics and transcriptional processes around *cis*-regulatory and coding regions [72]. This swapping mechanism, by which the entire or parts of the nucleosome are removed or replaced with new histones (canonical or variants) or their components, is known as "histone turnover" [71]. Thus, histone turnover plays an indispensable role in maintaining or altering the chemical nature, physical nature, and functional properties of nucleosomes, ultimately affecting regulatory processes like cell death.

Histone-modifying enzymes are one of the three important regulators of higher order chromatin conformation and nucleosome positioning, which eventually also control the regulation of gene expression. These histone-modifying enzymes are responsible for the covalent posttranslational modifications (PTM) of histone proteins. These PTM or "epigenetic marks" have been found in both flexible tails and globular domains of the canonical/core and linker histones. The extensive characterization of histone modifications has detected more than 550 posttranslational modifications, broadly categorized as epigenetic "writer," "reader," and "eraser." Epigenetic "writers" lay down specific modification on histone tail, which includes histone acetyltransferases (HATs), histone methyltransferases (HMT), protein arginine methyltransferases (PRMTs), and kinases. To read these marks and facilitate the binding of the transcriptional machinery, epigenetic "readers" are generated, including, for example, bromodomains, chromodomains, and tudor domains. The removal of particular marks is performed by epigenetic "erasers," such as histone deacetylases (HDACs), lysine demethylases (KDMs), and phosphatases [74, 75]. As proposed by C. David Allis and coworkers, the "histone code hypothesis" is based on these PTM through addition/removal or reading of covalent histone modifications to generate a remarkable diversity of combinatorial patterns, which govern biological specificity and downstream events [76]. An extension of this famous hypothesis is proposed to represent a broader "epigenetic code," which along with other factors also includes histone cassettes, binary switches, and the multivalency of effector-ligand binding reactions [77].

With detailed insight into the biological importance, a number of PTMs are being incorporated into the growing list of modifications. Usually, some PTMs like histone phosphorylation, acetylation, GlcNAcylation, palmitoylation, and methylation are "simple" modifications as they are reversible and binary, while some PTMs are more "complex" as they regulate important cellular responses like ubiquitination for protein degradation or poly-ADP-ribosylation for DNA damage responses [78]. The most widely studied PMT in cellular immunology is histone methylation. It occurs on the side chains of lysine and arginine residues. Like histone acetylation, it does not simply change charges of histones, but adds more methylation groups. It means lysine side chain of histones can be mono-, di-, or tri-methylated, whereas arginine can be mono- or di-methylated (symmetrically or asymmetrically) [79]. Thus, based on the presence of these histone modifications, functional feature of the regulatory elements, such as enhancers, can be represented. For example, H3K27ac (acetylation on 27th lysine residue of histone 3) with H3K4me1 (monomethylation on fourth lysine side chain of histone 3) marks represent active enhancer region, while only H3K4me1 represents poised enhancers [80]. Likewise, H3K4me3 marks active promoter and H3K27me3 marks repressed promoter [81]. Thus, as these examples illustrate, the presence of histone marks on different epigenetic features contributes to differential regulatory mechanisms that govern gene expression [82].

In the context of arginine and lysine methyltransferases, they structurally possess distinct catalytic site from DNA methyltransferases (DNMTs), but methyl group transfer is S-adenosylmethionine (SAM)-dependent. The histone methylations are considered more stable and static modifications. However, the emergence of histone demethylation played a very important role in the transcription control of gene expression and respective cellular processes. The jumonji protein JMJD6 represents an example involved in the reversal of arginine methylation on H3R2 and H4R3 [79]. However, findings on functional significant JMJD6-like modifications need to be reiterated in both healthy and disease conditions. Thus, other newly discovered PTM including lysine crotonylation (Kcr) and lysine 2-hydroxyisobutyrylation (Khib) are of interest for the functional characterization and specificity for epigenetic readers.

Posttranslational modifications: Nonhistone proteins

The nonhistone proteins are indispensable components of epigenetic mechanisms. Posttranslational modifications (PTMs) of nonhistone proteins emerged as prevalent and pivotal to many cellular processes such as apoptosis, metabolism, signal transduction, and inflammation in addition to the transcriptional control of gene regulation. More than 200 types of PTMs have been reported for nonhistone proteins, defining classes of "writers, readers, and erasers" as similar to histone proteins. Among other PTMs, acetylation, deacetylation, methylation has been studied. For instance, approximately 4000 lysine and arginine methylation sites for nonhistone proteins have been reported, which also share cross talk of individual modifications [83]. The importance of PTM of nonhistone proteins and its cross talk with histone proteins can be realized by acetylation/deacetylation of NF-kB (nuclear factor kB), a key transcription factor involved in the several cellular processes such as cell survival/death. It is found that acetylation of NK-kB (and its inhibition by SIRT1 histone deacetylases) triggers selective binding to IL-8 gene promoter and ultimately enhances transcriptional activity. Another example that illustrates the importance of deacetylation/acetylation of nonhistones in the inflammatory process is the glucocorticoid receptors (GR). Upon ligand binding, GR undergoes heavy acetylation but for the gene repression, deacetylation is necessary by binding to NF-kB [84].

Advancing technology and use of interdisciplinary approaches

Over the last decades, the biggest advancement in biological and clinical research is the development of sequencing technologies addressing questions from bulk-level to single immune cell [85]. The use of cutting-edge technologies like -omics has enabled the community to elucidate the comprehensive but high-resolution map of (*epi*)genetic mechanisms, often referred to as systems immunology. Certainly, the deeper understanding of epigenetic mechanisms has been equally important for diverse model organisms such as plants, insects, and so on [86]. Though epigenomics is only one of the several rapidly

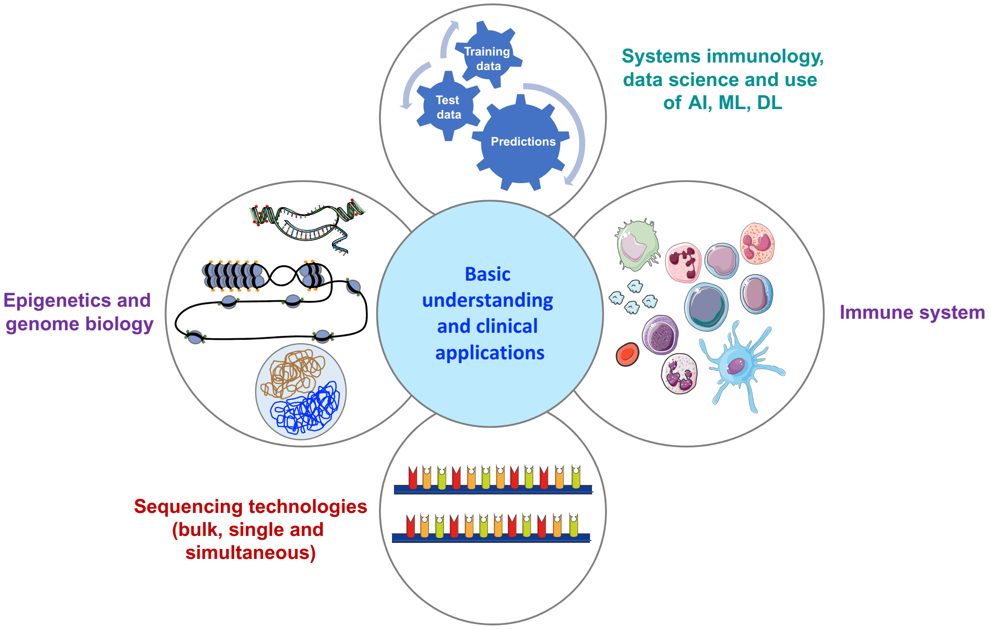

Training data
Test data
Predictions
Systems immunology, data science and use of AI, ML, DL
Epigenetics and genome biology
Basic understanding and clinical applications
Immune system
Sequencing technologies (bulk, single and simultaneous)

Fig. 5 The immune system and *epi*genetics at the interface of cutting-edge technologies. The immune system and epigenetic mechanisms are now being studied using different approaches. The cutting-edge technologies include the use of sequencing methods to understand the single (epi)genomic process and/or simultaneously from the same cell. The data generated from these methods are genome-scale and would be a good fit for data science. Data science relies on various approaches such as AI, ML, and DL. Thus, the combined knowledge of epigenetics and immunology together with novel technologies and computing methods would help to understand basic biology and apply this knowledge to the clinics. Abbreviations: *AI*, artificial intelligence; *DL*, deep learning; *ML*, machine learning.

growing -omics technologies, the main challenge for the next generation of scientists is to manage overwhelmingly generated "big data" [87, 88] and integrate these -omics layers, opening path to new field of scientific research famously known as "integrative -omics" or "multi-omics" [89]. To add more data, single-cell epigenomics is also heading toward such integrative approaches [90]. However, to rescue such mayhem of generated data, due to both addressing outstanding biological questions and use for personalized medicine/population-wide prediction studies of diseases, novel yet old approaches like machine learning have gained increasing interest (Fig. 5). Thus, big data science will be a promising aspect of future immunology research.

References

[1] Weaver C, Murphy KM. Janeway's immunobiology. 9th ed; 2017.
[2] Seddon B, Yates AJ. The natural history of naive T cells from birth to maturity. Immunol Rev 2018;285(1):218–32.
[3] Passos GA, et al. Update on Aire and thymic negative selection. Immunology 2018;153(1):10–20.
[4] Attias M, Al-Aubodah T, Piccirillo CA. Mechanisms of human FoxP3(+) Treg cell development and function in health and disease. Clin Exp Immunol 2019;197(1):36–51.
[5] Huehn J, Beyer M. Epigenetic and transcriptional control of Foxp3+ regulatory T cells. Semin Immunol 2015;27(1):10–8.
[6] Akkaya M, Kwak K, Pierce SK. B cell memory: building two walls of protection against pathogens. Nat Rev Immunol 2019;.
[7] Zhang Y, Good-Jacobson KL. Epigenetic regulation of B cell fate and function during an immune response. Immunol Rev 2019;288(1):75–84.
[8] Prager I, Watzl C. Mechanisms of natural killer cell-mediated cellular cytotoxicity. J Leukoc Biol 2019;105(6):1319–29.
[9] Vitale M, et al. An historical overview: the discovery of how NK cells can kill enemies, recruit defense troops, and more. Front Immunol 2019;10:1415.
[10] Cherrier DE, Serafini N, Di Santo JP. Innate lymphoid cell development: a T cell perspective. Immunity 2018;48(6):1091–103.
[11] Vivier E, et al. Innate lymphoid cells: 10 years on. Cell 2018;174(5):1054–66.
[12] Fruitwala S, El-Naccache DW, Chang TL. Multifaceted immune functions of human defensins and underlying mechanisms. Semin Cell Dev Biol 2019;88:163–72.
[13] Vidya MK, et al. Toll-like receptors: significance, ligands, signaling pathways, and functions in mammals. Int Rev Immunol 2018;37(1):20–36.
[14] Eisenbarth SC. Dendritic cell subsets in T cell programming: location dictates function. Nat Rev Immunol 2019;19(2):89–103.
[15] Embgenbroich M, Burgdorf S. Current concepts of antigen cross-presentation. Front Immunol 2018;9:1643.
[16] Brandes M, et al. Cross-presenting human gammadelta T cells induce robust CD8+ alphabeta T cell responses. Proc Natl Acad Sci USA 2009;106(7):2307–12.
[17] Iberg CA, Jones A, Hawiger D. Dendritic cells as inducers of peripheral tolerance. Trends Immunol 2017;38(11):793–804.
[18] Croft M, Dubey C. Accessory molecule and costimulation requirements for CD4 T cell response. Crit Rev Immunol 2017;37(2–6):261–90.
[19] Bashir B, Wilson MA. Novel immunotherapy combinations. Curr Oncol Rep 2019;21(11):96.
[20] Bhat J, et al. Stochastics of cellular differentiation explained by epigenetics: the case of T-cell differentiation and functional plasticity. Scand J Immunol 2017;86(4):184–95.

[21] Schmidl C, et al. Epigenetic mechanisms regulating T-cell responses. J Allergy Clin Immunol 2018;142(3): 728–43.
[22] Girardi M, et al. Regulation of cutaneous malignancy by gammadelta T cells. Science 2001;294 (5542):605–9.
[23] Hayday AC. Gammadelta T cell update: adaptate orchestrators of immune surveillance. J Immunol 2019;203(2):311–20.
[24] Boutin L, Scotet E. Towards deciphering the hidden mechanisms that contribute to the antigenic activation process of human Vgamma9Vdelta2 T cells. Front Immunol 2018;9:828.
[25] Chitadze G, et al. The ambiguous role of gammadelta T lymphocytes in antitumor immunity. Trends Immunol 2017;38(9):668–78.
[26] Schmolka N, et al. Epigenetic and transcriptional regulation of gammadelta T cell differentiation: programming cells for responses in time and space. Semin Immunol 2015;27(1):19–25.
[27] Bhat J, Kabelitz D. Gammadelta T cells and epigenetic drugs: a useful merger in cancer immunotherapy? Oncoimmunology 2015;4(6):e1006088.
[28] Bhoj VG, et al. Persistence of long-lived plasma cells and humoral immunity in individuals responding to CD19-directed CAR T-cell therapy. Blood 2016;128(3):360–70.
[29] Thai LH, et al. BAFF and CD4(+) T cells are major survival factors for long-lived splenic plasma cells in a B-cell-depletion context. Blood 2018;131(14):1545–55.
[30] Caccamo N, et al. Atypical human effector/memory CD4(+) T cells with a naive-like phenotype. Front Immunol 2018;9:2832.
[31] Nguyen QP, et al. Origins of CD4(+) circulating and tissue-resident memory T-cells. Immunology 2019;157(1):3–12.
[32] Pahl JHW, Cerwenka A, Ni J. Memory-like NK cells: Remembering a previous activation by cytokines and NK cell receptors. Front Immunol 2018;9:2796.
[33] Beaulieu AM. Memory responses by natural killer cells. J Leukoc Biol 2018;104(6):1087–96.
[34] Hammer Q, et al. Peptide-specific recognition of human cytomegalovirus strains controls adaptive natural killer cells. Nat Immunol 2018;19(5):453–63.
[35] Netea MG, et al. Trained immunity: a program of innate immune memory in health and disease. Science 2016;352(6284):aaf1098.
[36] Netea MG, et al. Innate and adaptive immune memory: an evolutionary continuum in the Host's response to pathogens. Cell Host Microbe 2019;25(1):13–26.
[37] Natoli G, Ostuni R. Adaptation and memory in immune responses. Nat Immunol 2019;20(7):783–92.
[38] Waddington CH. An introduction to modern genetics. Proc R Entomol Soc Lond A Gen Entomol 1939;14(4–6):82.
[39] Wu C, Morris JR. Genes, genetics, and epigenetics: a correspondence. Science 2001;293 (5532):1103–5.
[40] Goldberg AD, Allis CD, Bernstein E. Epigenetics: a landscape takes shape. Cell 2007;128(4):635–8.
[41] Passarge E. Emil Heitz and the concept of heterochromatin: longitudinal chromosome differentiation was recognized fifty years ago. Am J Hum Genet 1979;31(2):106–15.
[42] Brown SW. Heterochromatin. Science 1966;151(3709):417–25.
[43] Maston GA, Evans SK, Green MR. Transcriptional regulatory elements in the human genome. Annu Rev Genomics Hum Genet 2006;7:29–59.
[44] Risca VI, Greenleaf WJ. Unraveling the 3D genome: genomics tools for multiscale exploration. Trends Genet 2015;31(7):357–72.
[45] Leoni C, et al. Epigenetics of T lymphocytes in health and disease. Swiss Med Wkly 2015;145:w14191.
[46] Kornberg RD. Chromatin structure: a repeating unit of histones and DNA. Science 1974;184 (4139):868–71.
[47] Annunziato AT. DNA Packaging: Nucleosomes and Chromatin. Nat Educ 2008;1(1):26.
[48] Ramachandran S, Henikoff S. Nucleosome dynamics during chromatin remodeling in vivo. Nucleus 2016;7(1):20–6.
[49] Vermunt MW, Zhang D, Blobel GA. The interdependence of gene-regulatory elements and the 3D genome. J Cell Biol 2019;218(1):12–26.

[50] Kloetgen A, et al. 3D chromosomal landscapes in hematopoiesis and immunity. Trends Immunol 2019;40(9):809–24.

[51] Rao SSP, et al. Cohesin loss eliminates all loop domains. Cell 2017;171(2):305–20 [e24].

[52] Rowley MJ, Corces VG. Organizational principles of 3D genome architecture. Nat Rev Genet 2018;19(12):789–800.

[53] Stadhouders R, Filion GJ, Graf T. Transcription factors and 3D genome conformation in cell-fate decisions. Nature 2019;569(7756):345–54.

[54] van Schoonhoven A, et al. 3D genome organization during lymphocyte development and activation. Brief Funct Genomics 2019, elz030.

[55] Haarhuis JH, Rowland BD. Cohesin: building loops, but not compartments. EMBO J 2017;36 (24):3549–51.

[56] Chowdhury D, Novina CD. Potential roles for short RNAs in lymphocytes. Immunol Cell Biol 2005;83(3):201–10.

[57] Nguyen ML, et al. Transcriptional enhancers in the regulation of T cell differentiation. Front Immunol 2015;6:462.

[58] Araki Y, et al. Genome-wide analysis of histone methylation reveals chromatin state-based regulation of gene transcription and function of memory CD8+ T cells. Immunity 2009;30(6):912–25.

[59] Jonkers I, Lis JT. Getting up to speed with transcription elongation by RNA polymerase II. Nat Rev Mol Cell Biol 2015;16(3):167–77.

[60] Zorca CE, et al. Myosin VI regulates gene pairing and transcriptional pause release in T cells. Proc Natl Acad Sci USA 2015;112(13):E1587–93.

[61] Nie Z, et al. C-Myc is a universal amplifier of expressed genes in lymphocytes and embryonic stem cells. Cell 2012;151(1):68–79.

[62] Tullai JW, et al. Immediate-early and delayed primary response genes are distinct in function and genomic architecture. J Biol Chem 2007;282(33):23981–95.

[63] Derrien T, et al. The GENCODE v7 catalog of human long noncoding RNAs: analysis of their gene structure, evolution, and expression. Genome Res 2012;22(9):1775–89.

[64] Ranzani V, et al. The long intergenic noncoding RNA landscape of human lymphocytes highlights the regulation of T cell differentiation by linc-MAF-4. Nat Immunol 2015;16(3):318–25.

[65] Kowalczyk MS, Higgs DR, Gingeras TR. Molecular biology: RNA discrimination. Nature 2012;482 (7385):310–1.

[66] Yang XJ, Seto E. HATs and HDACs: from structure, function and regulation to novel strategies for therapy and prevention. Oncogene 2007;26(37):5310–8.

[67] Quinn JJ, Chang HY. Unique features of long non-coding RNA biogenesis and function. Nat Rev Genet 2016;17(1):47–62.

[68] Chen T, et al. M(6)a RNA methylation is regulated by microRNAs and promotes reprogramming to pluripotency. Cell Stem Cell 2015;16(3):289–301.

[69] Pai AA, et al. The contribution of RNA decay quantitative trait loci to inter-individual variation in steady-state gene expression levels. PLoS Genet 2012;8(10). e1003000.

[70] Petty E, Pillus L. Balancing chromatin remodeling and histone modifications in transcription. Trends Genet 2013;29(11):621–9.

[71] Venkatesh S, Workman JL. Histone exchange, chromatin structure and the regulation of transcription. Nat Rev Mol Cell Biol 2015;16(3):178–89.

[72] Weber CM, Henikoff S. Histone variants: dynamic punctuation in transcription. Genes Dev 2014;28(7): 672–82.

[73] Talbert PB, Henikoff S. Histone variants on the move: substrates for chromatin dynamics. Nat Rev Mol Cell Biol 2017;18(2):115–26.

[74] Falkenberg KJ, Johnstone RW. Histone deacetylases and their inhibitors in cancer, neurological diseases and immune disorders. Nat Rev Drug Discov 2014;13(9):673–91.

[75] Andrews FH, Strahl BD, Kutateladze TG. Insights into newly discovered marks and readers of epigenetic information. Nat Chem Biol 2016;12(9):662–8.

[76] Strahl BD, Allis CD. The language of covalent histone modifications. Nature 2000;403(6765):41–5.

[77] Allis CD, Jenuwein T. The molecular hallmarks of epigenetic control. Nat Rev Genet 2016;17(8): 487–500.
[78] Prabakaran S, et al. Post-translational modification: nature's escape from genetic imprisonment and the basis for dynamic information encoding. Wiley Interdiscip Rev Syst Biol Med 2012;4(6):565–83.
[79] Bannister AJ, Kouzarides T. Regulation of chromatin by histone modifications. Cell Res 2011;21(3): 381–95.
[80] Creyghton MP, et al. Histone H3K27ac separates active from poised enhancers and predicts developmental state. Proc Natl Acad Sci USA 2010;107(50):21931–6.
[81] Bernstein BE, et al. A bivalent chromatin structure marks key developmental genes in embryonic stem cells. Cell 2006;125(2):315–26.
[82] Pekowska A, et al. A unique H3K4me2 profile marks tissue-specific gene regulation. Genome Res 2010;20(11):1493–502.
[83] Biggar KK, Li SS. Non-histone protein methylation as a regulator of cellular signalling and function. Nat Rev Mol Cell Biol 2015;16(1):5–17.
[84] Ito K. Impact of post-translational modifications of proteins on the inflammatory process. Biochem Soc Trans 2007;35(Pt 2):281–3.
[85] Svensson V, Vento-Tormo R, Teichmann SA. Exponential scaling of single-cell RNA-seq in the past decade. Nat Protoc 2018;13(4):599–604.
[86] Gutierrez C. 25 years of cell cycle research: What's ahead? Trends Plant Sci 2016;21(10):823–33.
[87] Kahn SD. On the future of genomic data. Science 2011;331(6018):728–9.
[88] Stephens ZD, et al. Big data: astronomical or genomical? PLoS Biol 2015;13(7). e1002195.
[89] Misra BB, et al. Integrated omics: tools, advances, and future approaches. J Mol Endocrinol 2018;62(1): R21–45.
[90] Hu Y, et al. Single cell multi-omics technology: methodology and application. Front Cell Dev Biol 2018;6:28.

This page intentionally left blank

CHAPTER 2

Plant epigenetics and the ‘intelligent’ priming system to combat biotic stress

Hemlata Kotkar[a], **Ashok Giri**[b]
[a]Department of Botany, Savitribai Phule Pune University, Pune, India
[b]Biochemical Sciences Division, CSIR-National Chemical Laboratory, Pune, India

Contents

Introduction 25
Plant DNA methylation 26
De novo DNA methylation 27
Maintenance DNA methylation 28
Histone modifications 29
RNA-associated silencing 29
Epigenetics of plant microbe interactions 29
Epigenetics of plant insect interactions 30
Epigenetics of immune system and memory in plants 32
Plant epigenetics: Model plants and application in agriculture 33
Somatic embryogenesis 34
Heterosis 34
Conclusions 35
Acknowledgments 36
References 36

Introduction

Coined in 1942 by Conrad H. Waddington, the term “epigenetics” is derived by combining “epigenesis” and “genetics.” Epigenetic changes are heritable modifications in the genetic material such as adjustments in the pattern of DNA methylation and histones. A number of biotic and abiotic factors are known to influence epigenetic mechanisms in plants (Fig. 1).

These mechanisms do not affect the DNA sequence but can reversibly modify the expression of genes [1]. In plants, DNA methylation errors can lead to pleiotropic morphological defects. Methylation of DNA is its covalent modification and is common across many genera. Amongst these, plants are model study systems and have been reported to be dependent on epigenetic factors for growth and development. Such factors include the chromatin remodeling factor DDM1 (DECREASE IN DNA METHYLATION 1), the CG DNA methylation maintenance factor VIM1 (VARIANT IN METHYLATION 1)/UHRF1 (ubiquitin-like plant homeodomain and RING finger

Epigenetics of the Immune System
https://doi.org/10.1016/B978-0-12-817964-2.00002-2
© 2020 Elsevier Inc.
All rights reserved.

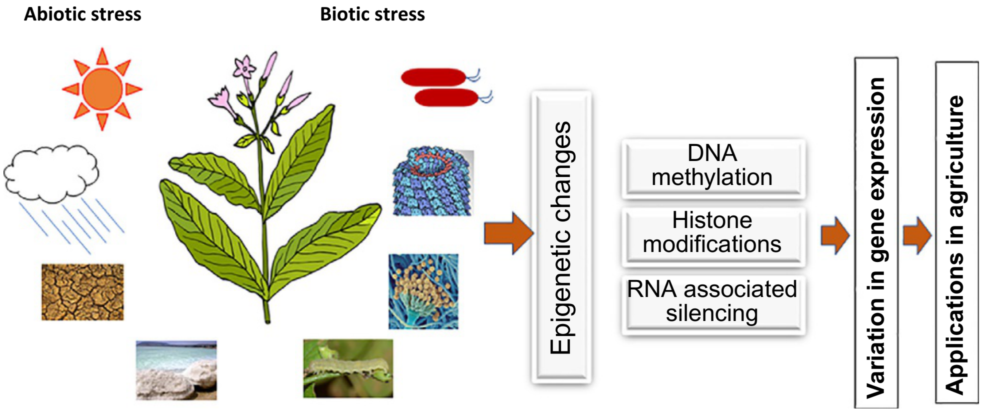


Fig. 1 ***Biotic stresses on plants influence epigenetic mechanisms.*** Plants are affected by abiotic stresses such as temperature, weather conditions, water availability, salinity and biotic stresses such as bacterial, viral, fungal and insect infestations. Epigenetic changes are mediated through DNA methylation, histone modifications and RNA associated silencing leading to variations in gene expression. These mechanisms have been understood and tapped for betterment for agriculture.

domain 1), the base-excision DNA demethylase DEMETER (DME glycosylase), and silencing components via small RNAs [2]. Although plants and animals show huge morphological differences and are evolutionarily separated, they have been known to share the same epigenetic mechanisms. Much of the epigenetic evidence is provided by evolutionary studies in plants that are more traceable as compared to animals. Using model systems, there is enough evidence that the plant phenotype is controlled not only by genetic but also by epigenetic factors. However, such mechanisms in both animals and plants can be correlated and occur due to three different systems that exist within cells to silence genes: DNA methylation, histone modification, and RNA-associated silencing.

Plant DNA methylation

DNA methylation is a highly specific chemical process of addition of a methyl group to DNA (Fig. 2A). This takes place when a cytosine (C) is placed near guanine (G) which has a phosphate (p) linked, called a CpG site. These sites are methylated by DNA methyltransferases, altering the structure of DNA. Cytosine methylation is a modification of DNA that has been reported to be stable and the one that can be inherited. Methylome sequencing studies provide an estimate of methylation. DNA methylation that is genome-wide can occur at levels of 24%, 6.7%, and 1.7% for CG, CHG, and CHH contexts (where H = A, T, or C), respectively, in the model plant, *Arabidopsis thaliana* L. [2]. Methylation reactions are highly important and provide epigenetic control in the genome. They also regulate coding and noncoding elements. This is reflected by the fact

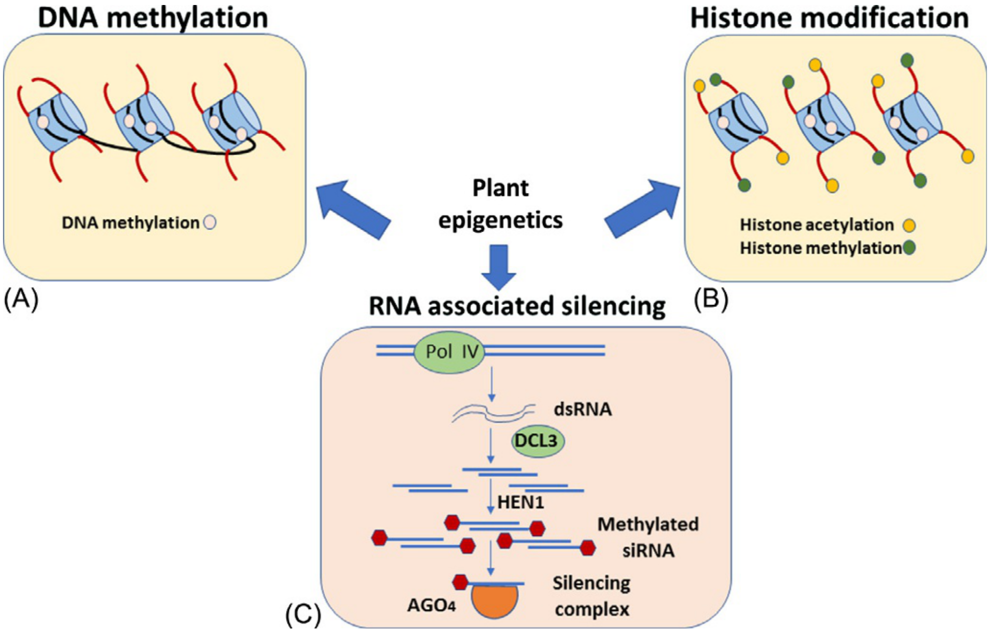

DNA methylation
DNA methylation ○
(A)
Plant epigenetics
Histone modification
Histone acetylation ●
Histone methylation ●
(B)
RNA associated silencing
Pol IV
dsRNA
DCL3
HEN1
Methylated siRNA
Silencing complex
AGO4
(C)

Fig. 2 ***Model for cell systems leading to epigenetic mechanisms in plants.*** Epigenetic mechanisms occur due to three important systems existing in cells: (A) DNA methylation wherein an addition of a methyl group to the DNA molecule takes place. (B) Histone modification including acetylation/methylation and (C) RNA associated silencing: Single stranded RNA (ssRNA) is generated by Polymerase IV (Pol IV). Later, dsRNA is generated from ssRNA by RNA-DEPENDENT RNA POLYMERASE 2 (RDR2). This dsRNA is processed by DICER-LIKE 3 (DCL3) into small interfering RNAs (siRNAs). Hua Enhancer1 (HEN1) methylates siRNAs at their 3′ ends. Methylated RNA combines with AGO4 (Argonaute protein 4) that contains RNA-induced silencing complex (RISC).

that several important processes in plants such as response to a pathogen, stability of the genome, heterosis, imprinting, transposable element regulation, and expression of genes are affected by DNA methylation [3]. DNA methylation is therefore considered to be the key factor in plant adaptation and evolution. It occurs predominantly on transposons as well as other repetitive DNA elements. Methylation of DNA has been studied extensively and categorized into two major types, viz.: (i) de novo DNA methylation and (ii) maintenance DNA methylation. In the following sections, we will focus on them with relevant illustrations.

De novo DNA methylation

All eukaryotic cells exhibit a mechanism by which foreign RNA molecules can be degraded as a part of the defense strategy. This mechanism is called the RNA interference (RNAi) pathway and protects plants against RNA viruses. The presence of a free pool of RNA molecules that are double-stranded triggers the mechanism of RNAi. RNAi involves a protein complex consisting of RNA nuclease and RNA helicase that cleave double-stranded RNA into small fragments. Thus, the RNAi pathway generates small interfering RNAs that are 24 nucleotide (nt) long and are associated with the enzyme. Other RNA molecules containing complementary strands are targeted. Interestingly, during the process of development of plants, these small RNAs play an important role in targeting genomic DNA sequences for methylation of cytosine. As discussed in the preceding section, cytosine methylation is a stable form of modification of DNA. This phenomenon in which RNA is involved is called the RNA-directed DNA methylation pathway (RdDM), first described by Wassenegger et al. [4].

De novo DNA methylation is involved in many biological processes like plant development, responses to stress that might be biotic/abiotic, along with maintaining the inert state of transposons/repetitive sequences in *Arabidopsis* [2]. Apart from the complete RNAi machinery, de novo DNA methylation in plants depends on two RNA polymerases that are plant specific, Pol IV and Pol V. A few other proteins and chromatin remodeling factors are also required. Pol IV is RNA-DEPENDENT RNA POLYMERASE 2 (RDR2) that is essential for generating siRNA required for DNA methylation. In a nutshell, for the accumulation of siRNA, Domains Rearranged Methyltransferase 2 (DRM2), ARGONAUTE 4 (AGO4), and Pol V are needed. In the course of plant development, transposon reactivation during male gametogenesis and genome-wide losses of DNA methylation in the process of female gametogenesis can occur [2]. Therefore, once DNA methylation is established, it has to be stably maintained to ensure cell identity.

Maintenance DNA methylation

In maintenance DNA methylation, MET1 (Methyltransferase 1) maintains CG methylation. Two DRM enzymes, Chromomethylase 2 (CMT2) and CMT3 maintain non-CG contexts. Plant-specific CMT3 recognizes H3K9 methylation and prefers CHG sites [5, 6]. Loss of CMT3 or the histone methyltransferase responsible for H3K9 demethylation, known as SUPPRESSOR OF VARIEGATION 3-9 HOMOLOGUE 4 (SUVH4, also known as KYP), leads to a decrease in DNA methylation [5]. Any cross talk between CMT3 and SUVH4 in the form of direct protein interaction for maintenance of CHG methylation is not reported in plants. However, CG methylation in plants and animals is maintained in a similar fashion. As discussed earlier, DDM1 is a chromatin remodeling factor involved in the maintenance of DNA methylation [7]. MET1 maintains CG methylation in coding regions of almost one-third of genes in *A. thaliana* [8]. Proteins with SRA (SET and Ring Finger associated) domains are required by RdDM for the maintenance of CG and CHG methylation. SUVH9 and SUVH2 are rich in SRA domains. SUVH9 selectively binds CHH while SUVH2 binds to CG for methylation.

When there is a lack of active maintenance of DNA methylation, demethylation occurs. Catalytic removal of 5-methylcytosine leads to active DNA demethylation [9].

To perform this task, plants have been reported to encode a family of glycosylases in combination with the base excision repair (BER) pathway. Glycosylases can actively remove DNA methylation. DEMETER (DME), REPRESSOR OF SILENCING 1 (ROS1), DEMETER-LIKE 2(DML2), and DML3 are DNA glycosylases reported from *Arabidopsis* [2]. Locus-specific loss of DNA methylation can be induced when glycosylases act in a developmentally programmed fashion. Despite having the same substrates, the roles of these enzymes differ viz. DME is important during gametogenesis while others function in vegetative tissues.

Histone modifications

Covalent modifications, including acetylation, methylation (Fig. 2B), phosphorylation, ubiquitination, and sumoylation, can occur in histone proteins. Chromatin structure and function can be affected by incorporation of histone variants and relocation of nucleosomes. DNA methylation can be influenced by histone modifications in plants. Methylation is more at repeat sequences in H3K9 [10]. Protein homologs (15) of H3K9 methyltransferase SU(VAR)39 have been reported in *Arabidopsis* [11].

RNA-associated silencing

Genes can be turned off by RNA as antisense transcripts or noncoding RNAs by a phenomenon known as RNA interference. In other words, RNA is able to silence or negatively regulate expression of a gene by a mechanism called RNA interference. It is also known as posttranscriptional gene silencing (PTGS).

Small interfering RNAs (siRNAs) affect gene expression by formation of heterochromatin or by histone modifications and DNA methylation. siRNAs are highly responsive to stress and are therefore reported as the most stress-rearranged molecules. Importantly, they are known to carry memory of stress that has been experienced in an earlier situation to the next situation [12–14]. Fig. 2C provides a quick overview of the mechanism involved in RNA-associated silencing. Briefly, it employs Pol IV specific to plants to form specific single-stranded RNAs from dsRNAs. Also, Dicer-like 3 (DCL3) are involved that cleave dsRNAs into 20 to 24-nt siRNAs. Hua Enhancer1 (HEN1) methylates siRNAs at their 3′ ends. One such methylated RNA combines with AGO4 (Argonaute protein 4) that contains RNA-induced silencing complex (RISC).

Epigenetics of plant microbe interactions

Plants, being rich sources of nutrients, are under continuous threat of being attacked by enemies singly or in multiples. Plants lack an adaptive immune system like animals to combat microbial pathogens. However, they use physical barriers like the waxy cuticle to protect themselves. In order to face unfavorable conditions, plants have brought subtle changes to prime their immune system comprising constitutive and systemic defenses. Plant-pathogen interaction and plant-plant communication are dependent on the cross talk between key signaling chemical entities such as (i) jasmonic acid (JA), (ii) salicylic acid (SA), and (iii) reactive oxygen species. Plants pool danger signals just near the wounds and the pathogen invasion area. Apart from local defense responses, a microbial attack can trigger defense responses systemically. Systemic resistance involves two forms: systemic acquired resistance (SAR) and induced systemic resistance (ISR). The first form results from an earlier acquaintance of a plant to a pathogen that can be responsible for enduring long-term immunity after further attacks by the pathogen [15, 16]. This mechanism involves two essential components, viz., SA, a plant hormone, and NPR1 (non-expressor of pathogenesis-related genes 1), the downstream signaling protein [17, 18]. In SAR, transcription of all genes related to SA is activated. This essentially includes genes encoding for antimicrobial pathogenesis-related proteins (PR) [19]. In turn, high levels of SA are also responsible for target gene chromatin modification [20]. For instance, at the PR-1 promoter, levels of H3 and H4 acetylation and H3K4 methylation are enhanced. Similarly, as discussed in the preceding section, WRKY genes encode transcription factors that can be "turned on" whenever the plant is under threat of pathogen infection or SA treatment is induced [21]. In similar experiments, it was found that recombination frequency in *Nicotiana tabacum* upon infection with Tobacco Mosaic Virus (TMV) was enhanced. This eventually led to resistance in *N. tabacum* against TMV for two subsequent generations [22].

Epigenetics of plant insect interactions

Plants are highly sessile organisms in nature and their performance is known to be affected by abiotic factors such as temperature, light, water, and nutrient availability. However, plants also have their own biotic environment that plays a significant role [12] (Table 1).

For example, plant-associated biota such as pollinators, rhizobia, and mycorrhiza are highly beneficial while pathogens and herbivores can be detrimental. Using 'intelligent' strategies, plants have to optimize their signals to attract pollinators and deter herbivores. During this process, plants need mechanisms to keep a check on the insect community, identify, and quickly respond to potential threats [27]. DNA methylations mediate such responses and have been known to be heritable. Likewise, CG methylation is around 80% in plants and is organized in a tissue-specific manner [30]. Plants have evolved both chemical and physical barriers that function as defense strategies against herbivory [31]. Classically what is called as the evolutionary arms race between plants and animals, it has led to progression of plant mechanisms to ward off insects and on the other hand developed insects to inactivate plant chemicals [32]. So, the question that arises is, how does epigenetics come into play? This happens whenever herbivory sets in to induce

Table 1 Control of plant gene expression in response to adverse biotic factors.

Biotic factor	Plant	Parameters affected	Citation
Herbivory	*Impatiens capensis*	Flowering, plant height, biomass	Steets and Ashman [23]
	Raphanus raphanistrum	Seed mass, seedling growth, leaf trichome density	Agrawal et al. [24], Agrawal [25, 26]
	Brassica rapa	Variation in floral scent	Kellenberger et al. [27]
Simulated Herbivory	*Mimulus guttatus*	Leaf trichome density	Holeski [28]
Viral infection			
Tobacco mosaic virus	*Nicotiana tabacum*	Pathogen resistance, homologous recombination frequency	Kathiria et al. [22]
Bacterial infection			
Pseudomonas syringae	*Arabidopsis thaliana*	DNA methylation	Dowen et al. [13]
Xanthomonas oryzae	*Oryza sativa*	Increased resistance to blight	Alonso et al. [12]
Fungal infection			
Alternaria brassicicola *Botrytis cinerea*	*Arabidopsis thaliana*	Resistance to fungal pathogens	Zhou et al. [29]

Biotic factors such as herbivory, viral, and bacterial infection are responsible for affecting plant performance. Citations mentioned in the table provide details of plant parameters affected.

stress on the plant. This stress in turn triggers methylation changes in the defense-related genes. Several examples have been reported that provide a direct clue to epigenetic changes induced by herbivory. DNA methylation has been reported as a floral signal response to herbivory in yet another model crop plant species, *Brassica rapa* L. [27]. Methylation changes that were genome-wide were detected using a methylation-sensitive amplified length polymorphism (MSAP). These changes were observed in leaves as well as undamaged flowers upon foliar herbivory by the specialist butterfly, *Pieris brassicae* L. One clear indication of "transgenerational memory" was revealed when wild radish (*Raphanus raphanistrum*) resistant seedlings were produced by plants that were damaged by insects as compared to those that were undamaged [24].

It was observed that when a caterpillar chewed the first leaf, the density of trichomes increased from the third to the seventh leaf, preventing caterpillars from attacking new leaves. Increased trichome production was also seen in yellow monkey flower (*Mimulus guttatus*) fed by insects in the previous generation [28]. *Arabidopsis* and *Solanum lycopersicum* L. (tomato) plants allowed to be damaged by herbivory were more resistant when they encountered the same condition in their next generation [33]. In *Arabidopsis*, this phenomenon of transgenerational resistance to deal with herbivores has been well explained. It includes the priming of jasmonic acid (JA)-related defense responses and also depends on siRNA biogenesis as described earlier. However, such herbivory-induced changes can be detrimental and alter the response of plants with beneficial insects such as pollinators.

During feeding on plants, specialized insects adapt to their diets by developing strategies against toxic host components. On the other hand, when this adaptability is high, other nutritional diets might be overlooked. A slight increase in the efficiency of ingested (ECI) and digested (ECD) food utilization in spruce budworm (*Choristoneura fumiferana* Clem.) after two generations in low-quality food was observed. However, it remains to be determined whether this response is genetic or epigenetic [34].

Epigenetics of immune system and memory in plants

Plants cannot change locations but they can endure adverse environmental changes and retain memory through generations. Memory is usually used to define the ability of organisms to learn from their earlier experiences and is associated with higher organisms. It has been studied to a greater extent in animals which have brains and tissues to transmit action potentials to cope with stress. Plants have the ability to "sense," store, and also recall such earlier stress events. Information that is continuously perceived is then used to modify responses if they face new challenges. Environment-induced chromatin modifications at various responsive genes contribute principally to environmental memories in plants [35]. With the possession of such intricate systems, plants in a way similar to animals can handle stresses efficiently. It has been proposed by some plant scientists to consider plants as intelligent organisms [36] in view of the fact that "plant neurobiology" can throw light on several mechanisms of plant behavior [37]. It has been suggested that plants have nerves, a brain-equivalent in the roots and an intelligence.

However, the control of gene expression is important to face certain unavoidable circumstances, for example, different biotic stresses. The evolution of such resource-saving stress fighters by plants is inducible and based upon a past experience. The strategy that allows producing an effective response without a delay is called "priming." In plants, such 'intelligent' priming systems and somatic memory to pathogens have been reported [38, 39]. Epigenetic environmental memories are those marks that are maintained in the absence of inducers. This results in acquired "memorization" of environmental experiences in plants. Such memory in plants is often displayed as increased or hypersensitive response to recurring stresses, leading to enhanced defenses against pathogens and insects or improved stress resistance and adaptation. In other words, plants are "primed" by the first encounter of a stress and can "memorize" past environmental experiences to be better prepared for a recurring event. Most environmental memories are relatively short term (mitotic) and only last for the life span of an individual (somatic memory), whereas some can be long term (meiotic) and transmitted to the next generation (transgenerational memory) [40]. However, any induced response or maintenance of primed transcription and epigenetic memory in plants directly relates to fitness costs. Plants rely on two forms of innate immunity called the basal or horizontal disease resistance and gene-based or vertical disease resistance. In newer terminology, these are referred to as PAMP (pathogen-associated molecular pattern)-triggered immunity (PTI) and effector-triggered immunity (ETI) [41].

In an experiment, silencing using RNAi in *Nicotiana attenuata* plants for the expression of two transcription factors: WRKY3 and WRKY6 was targeted. This was achieved by transformation with inverted-repeat (ir) fragments of endogenous genes. Differences in responses were recorded to multiple elicitations as compared to a single elicitation. Both ir-wrky3 and ir-wrky6 plants exhibit normal JA bursts. But they exhibited lower JA accumulation and trypsin protease inhibitors (TPIs) after multiple elicitation or after feeding by *Manduca sexta* neonates. This proved that WRKY-deficient plants are not able to utilize their past experiences for a defense response. However, it was noted that *N. attenuata* WRKY transcription factors that were induced by herbivory have a significant role in generating priming imprints in the form of JA or JA precursor pools implicating plant memory [42].

Plant epigenetics: Model plants and application in agriculture

In the agricultural ecosystem, plants are highly responsive to physiological or developmental changes whenever they sense different stresses. Floral development, time of flowering induction, and response to stresses can be altered due to adjustments. While studying such systems, the first instance of application of epigenetics dates back to the knowledge of "paramutation" that was an evidence for non-Mendelian epigenetic inheritance. It has been first reported in two model and commercially important crops, maize and tomato [43]. The expression of a single allele which could be either of maternal or paternal origin is called parental imprinting and was first observed in maize [43]. Later, these systems have been studied in humans and reported as underlying mechanisms for multiple genetic disorders [44].

Plants serve as model systems for the study of epigenetic mechanisms due to several reasons, viz.: (i) they can undergo mutagenesis by chemical or physical treatments, random insertion of transgenes, or mobilization of transposable elements, (ii) self-pollinated plants can be easily identified, (iii) the RNAi mechanism to knock down the expression of candidate genes is easily facilitated, and (iv) many epigenetic mutants can be tolerated without threatening the life of a plant. Plant epigenetics has been known to benefit agriculture as discussed briefly in the following sections.

Somatic embryogenesis

It is a common practice in plant tissue culture wherein differentiated somatic cells are reprogrammed to form somatic or "vegetative" embryos. These somatic embryos develop into plants without requiring fertilization. In such a population, genetically uniform clones are expected but phenotypic variability is usually present to some extent. Such a change is called "somaclonal variation," which is linked to epigenetic basis. These variations are useful in the field of plant breeding and selection of adaptive traits [45]. It has been reported that epigenetic signals can be diffusible and travel through plasmodesmata and the vascular system. Therefore, in grafted plants, they can be even transferred between roots to shoots [46].

Heterosis

Heterosis or hybrid vigor, first introduced by George Shull, is described as a phenomenon in which a hybrid progeny or F1 phenotype is superior when compared to any of its parental inbred lines. The term heterosis in general includes somatic hybrids and heterozygotes. In plants, heterosis is an important phenomenon because it is responsible for vigor in growth (higher biomass) and fitness (increased resistance to various biotic and abiotic stresses) [47]. Although molecular mechanisms underlying heterosis have not been understood clearly, three hypotheses of gene activity between two genomes have been explained: dominance, overdominance, and epistasis. Dominance heterosis suggests the effect of dominant alleles over recessive alleles. It means that an inbred line having equally better performance as compared to the F1 hybrid cultivar can be created. This is achieved by removing all deleterious alleles and/or pooling favorable alleles [47]. The overdominance hypothesis proposes that hybrid heterozygous genotypes are better as compared to any of the parental states [48]. Thus, the term epistasis links to interactions involving two or multiple genes originating from different parents aiming to develop an improved phenotype [49].

The molecular basis of heterosis has been detailed by modern "omics" tools such as transcriptomics, proteomics, metabolomics, and epigenomics (including DNA methylomes, small RNAomes, and genome-wide distribution of histone modifications). In the model plant, *A. thaliana*, loci responsible for phenotypic variation caused due to epigenetic regulation were identified. Epigenetic recombinant inbred lines, called epi-RILs between parents, with differences only in their epigenetic marks, such as hybrids between *met1* (methyltransferase 1) and wt (wild type) Col, or between *ddm1* and wt Col were generated [50–52]. Relatively complex traits, for example, time of flowering, primary root length, plant height, and biomass, have been noted in these epi-RIL populations [50, 51, 53]. Research in *A. thaliana* hybrids show different levels of heterosis in vegetative biomass when parental combinations were altered [7]. The heterosis phenotype can be observed within a few days of sowing wherein hybrids show an increased size of the cotyledon as compared to parents. When this is related to the photosynthetic efficiency, it is found to be the same in parents and hybrids. But, because of the larger leaves in hybrids, the total amount of photosynthesis is greater than in parents. Such increased total photosynthesis might contribute to the heterosis phenotype. Trait selection of plant size in two backcrosses provided substantial recovery of the genetic background to produce heterosis when crossed to another parental line [7].

The importance in understanding heterosis and implying its applications lies in the fact that there is immense pressure on increasing crop productivity to feed the ever-increasing population. Plant breeders have used hybrid vigor extensively in crops such as maize (*Zea mays*), rice (*Oryza sativa*), canola (*Brassica napus*), sorghum (*Sorghum bicolor*), and a number of vegetables. As an important example, in 1935, hybrid corn was less than 10% in the state of Iowa but increased to over 90% 4 years later. Hybrid corn was the main crop in the 1950s in the United States, and yield advanced from 1 ton/ha in 1930 to 4 ton/ha in 1960 and approximately 12 tons/ha in 2017 [54, 55]. Another example dates back to the 1970s when hybrid rice was developed in China. This rice was accepted as there was 10 to 20% yield advantage over inbred parental varieties, increasing the area under cultivation [56].

As compared to animals, plants have more tolerance to polyploidy, i.e., the multiplication of whole genome chromosome complement. Polyploid crops such as *Triticum aestivum* L., *Gossypium* species, *Solanum tuberosum* L., *Arachis hypogaea* L., *Saccharum officinarum* L., *Coffea arabica* L., and cultivars of the plant family *Brassicaceae* and *Nicotiana* species have been developed for various advantages, such as hybrid vigor or resistance to the effects of deleterious mutations, thereby allowing duplicated genes to potentially acquire beneficial mutations. Also, formation of polyploids is linked with important changes both genomic and epigenetic [57]. Thus, agriculture has advanced with the know-how of heterosis by breeding crops to improve their performance for applications directly related to benefit humankind [58].

Conclusions

Plants use different mechanisms for the control of DNA methylation. The genome and epigenome organization between animals and plants has been found to be similar in terms of genome size and complexity and euchromatin to heterochromatin ratios [43]. Plants have been used as model systems to study epigenetic changes. The importance of methylation of DNA and its implications are being extensively intrigued. The cross talk between the plant methylome, histone modification, and sRNAs provides tools for betterment of commercially important crops in the near future.

Acknowledgments

HMK is thankful for financial assistance by UGC-DSA to the Department of Botany, Savitribai Phule Pune University.

References

[1] Iwasaki M, Paszkowski J. Epigenetic memory in plants. EMBO J 2014;33(18):1987–98. https://doi.org/10.15252/embj.201488883.

[2] Law JA, Jacobsen SE. Establishing, maintaining and modifying DNA methylation patterns in plants and animals. Nat Rev Genet 2010;11(3):204–20. https://doi.org/10.1038/nrg2719.

[3] Vidalis A, Živković D, Wardenaar R, Roquis D, Tellier A, Johannes F. Methylome evolution in plants. Genome Biol 2016;17(1):1–14. https://doi.org/10.1186/s13059-016-1127-5.

[4] Wassenegger M, Heimes S, Riedel L, Sänger HL. RNA-directed de novo methylation of genomic sequences in plants. Cell 1994;76(3):567–76. https://doi.org/10.1016/0092-8674(94)90119-8.

[5] Jackson JP, Lindroth AM, Cao X, Jacobsen SE. Control of CpNpG DNA methylation by the KRYPTONITE histone H3 methyltransferase. Nature 2002;416:556–60. https://doi.org/10.1038/nature731.

[6] Lindroth A, Shultis D, Jasencakova Z, Fuchs J, Johnson L, Schubert D, et al. Dual histone H3 methylation marks at lysines 9 and 27 required for interaction with CHROMOMETHYLASE3. EMBO J 2004;23:4286–96.

[7] Kawanabe T, Ishikura S, Miyaji N, Sasaki T, Wu LM, et al. Role of DNA methylation in hybrid vigor in *Arabidopsis thaliana*. Proc Natl Acad Sci USA 2016;113(43):E6704–11. https://doi.org/10.1073/pnas.1613372113.

[8] Cokus SJ, Feng S, Zhang X, Chen Z, Merriman B, Haudenschild CD, et al. Shotgun bisulphite sequencing of the *Arabidopsis* genome reveals DNA methylation patterning. Nature 2008;452:215–9.

[9] Zhang H, Zhu J. Active DNA demethylation in plants and animals. Cold Spring Harb Symp Quant Biol 2012;77:161–73.

[10] Saze H, Tsugane K, Kanno T, Nishimura T. DNA methylation in plants: Relationship to small rnas and histone modifications, and functions in transposon inactivation. Plant Cell Physiol 2012;53(5):766–84. https://doi.org/10.1093/pcp/pcs008.

[11] Baumbusch LO. The *Arabidopsis thaliana* genome contains at least 29 active genes encoding SET domain proteins that can be assigned to four evolutionarily conserved classes. Nucleic Acids Res 2002;29(21):4319–33. https://doi.org/10.1093/nar/29.21.4319.

[12] Alonso C, Ramos-Cruz D, Becker C. The role of plant epigenetics in biotic interactions. New Phytol 2019;221(2):731–7. https://doi.org/10.1111/nph.15408.

[13] Dowen RH, Pelizzola M, Schmitz RJ, Lister R, Dowen JM, Nery JR, Ecker JR. Widespread dynamic DNA methylation in response to biotic stress. Proc Natl Acad Sci USA 2012;109(32):E2183–91. https://doi.org/10.1073/pnas.1209329109.

[14] Pieterse CMJ. Prime time for transgenerational defense. Plant Physiol 2012;158(2):545. https://doi.org/10.1104/pp.112.900430.

[15] Vlot AC, Klessig DF, Park SW. Systemic acquired resistance: the elusive signal(s). Curr Opin Plant Biol 2008;11(4):436–42. https://doi.org/10.1016/j.pbi.2008.05.003.

[16] Vlot A, Liu P, Cameron R, Park S, Yang Y, Kumar D, Zhou F, Padukkavidana T, Gustafsson C, Pichersky E, et al. Identification of likely orthologs of tobacco salicylic acid-binding protein 2 and their role in systemic acquired resistance in *Arabidopsis thaliana*. Plant J 2008;56:445–56.

[17] Durrant W, Dong X. Systemic acquired resistance. Annu Rev Phytopathol 2004;42:185–209.

[18] Loake G, Grant M. Salicylic acid in plant defence—the players and protagonists. Curr Opin Plant Biol 2007;10(5):466–72. https://doi.org/10.1016/j.pbi.2007.08.008.

[19] Ryals J, Neuenschwander U, Willits M, Molina A, Steiner H, Hunt M. Systemic acquired resistance. Plant Cell 1996;8:1809–19.

[20] Butterbrodt T, Thurow C, Gatz C. Chromatin immunoprecipitation analysis of the tobacco PR-1a- and the truncated CaMV 35S promoter reveals differences in salicylic acid-dependent TGA factor binding and histone acetylation. Plant Mol Biol 2006;61(4–5):665–74. https://doi.org/10.1007/s11103-006-0039-2.

[21] Asai T, Tena G, Plotnikova J, Willmann M, Chiu W, et al. MAP kinase signalling cascade in *Arabidopsis* innate immunity. Nature 2002;415:977–83.

[22] Kathiria P, Sidler C, Golubov A, Kalischuk M, Kawchuk LM, Kovalchuk I. Tobacco mosaic virus infection results in an increase in recombination frequency and resistance to viral, bacterial, and fungal pathogens in the progeny of infected tobacco plants. Plant Physiol 2010;153:1859–70.

[23] Steets J, Ashman T. Maternal effects of herbivory in *Impatiens capensis*. Int J Plant Sci 2010;171:509–18.

[24] Agrawal A, Laforsch C, Tollrian R. Transgenerational induction of defences in animals and plants. Nature 1999;401:60–3. https://doi.org/10.1038/43425.

[25] Agrawal A. Herbivory and maternal effects: mechanisms and consequences of transgenerational induced plant resistance. Ecology 2002;83:3408–15. https://doi.org/10.1890/0012-9658(2002)083[3408,HAMEMA]2.0.CO;2.

[26] Agrawal A. Phenotypic plasticity in the interactions and evolution of species. Science 2001;294(5541):321–6. https://doi.org/10.1126/science.1060701.11598291.

[27] Kellenberger RT, Schlüter PM, Schiestl FP. Herbivore-induced DNA demethylation changes floral signalling and attractiveness to pollinators in *Brassica rapa*. PLoS ONE 2016;11(11):1–17. https://doi.org/10.1371/journal.pone.0166646.

[28] Holeski L. Within and between generation phenotypic plasticity in trichome density of *Mimulus guttatus*. J Evol Biol 2007;20:2092–100.

[29] Zhou C, Zhang L, Duan J, Miki B, Wu K. HISTONE DEACETYLASE19 is involved in jasmonic acid and ethylene signalling of pathogen response in *Arabidopsis*. Plant Cell 2005;17:1196–204. https://doi.org/10.1105/tpc.104.028514.

[30] Widman N, Feng S, Jacobsen S, Pellegrini M. Epigenetic differences between shoots and roots in *Arabidopsis* reveals tissue-specific regulation. Epigenetics 2014;9:236–42.

[31] Karban R, Baldwin I. Induced responses to herbivory. Chicago: University of Chicago Press; 1997.

[32] Crava CM, Brütting C, Baldwin IT. Transcriptome profiling reveals differential gene expression of detoxification enzymes in a hemimetabolous tobacco pest after feeding on jasmonate-silenced *Nicotiana attenuata* plants. BMC Genomics 2016;17(1):1–15. https://doi.org/10.1186/s12864-016-3348-0.

[33] Rasmann S, De Vos M, Casteel CL, Tian D, Halitschke R, Sun J, Agrawal A, Felton G, Jander G. Herbivory in the previous generation primes plants for enhanced insect resistance. Plant Physiol 2012;158(2):854–63. https://doi.org/10.1104/pp.111.187831.

[34] Quezada-García R, Fuentealba Á, Bauce É. Phenotypic variation in food utilization in an outbreak insect herbivore. Insect Sci 2018;25(3):467–74. https://doi.org/10.1111/1744-7917.12419.

[35] He Y, Li Z. Epigenetic environmental memories in plants: establishment, maintenance, and reprogramming. Trends Genet 2018;34(11):856–66. https://doi.org/10.1016/j.tig.2018.07.006.

[36] Brenner ED, Stahlberg R, Mancuso S, Baluska F, Van Volk-Enburgh E. Response to Alpi et al.: plant neurobiology: the gain is more than the name. Trends Plant Sci 2007;12:285–6.

[37] Alpi A, Amrhein N, Bertl A, et al. Plant neurobiology: no brain, no gain? Trends Plant Sci 2007;12:135–6. https://doi.org/10.1111/j.1365-3040.2008.01862.x.

[38] Crisp P, Ganguly D, Eichten S, Borevitz J, Pogson B. Reconsidering plant memory: intersections between stress recovery, RNA turnover, and epigenetics. Sci Adv 2016;2:e1501340.

[39] Lamke J, Baurle I. Epigenetic and chromatin-based mechanisms in environmental stress adaptation and stress memory in plants. Genome Biol 2017;18:124.

[40] Cazzonelli CI, Millar T, Finnegan EJ, Pogson BJ. Promoting gene expression in plants by permissive histone lysine methylation. Plant Signal Behav 2009;4(6):484–8. https://doi.org/10.4161/psb.4.6.8316.

[41] Chisholm ST, Coaker G, Day B, Staskawicz BJ. Host-microbe interactions: shaping the evolution of the plant immune response. Cell 2006;124:803–14.

[42] Galis I, Gaquerel E, Pandey SP, Baldwin IT. Molecular mechanisms underlying plant memory in JA-mediated defence responses. Plant Cell Environ 2009;32(6):617–27. https://doi.org/10.1111/j.1365-3040.2008.01862.x.

[43] Pikaard C, Mittelsten S. Epigenetic regulation in plants. Cold Spring Harb Perspect Biol 2014;6: a019315. https://doi.org/10.1101/cshperspect.a019315.

[44] Zoghbi H, Beaudet A. Epigenetics and human disease. Cold Spring Harb Perspect Biol 2014;8: a019497. https://doi.org/10.1101/cshperspect.a019497.

[45] Miguel C, Marum L. An epigenetic view of plant cells cultured *in vitro*: soma-clonal variation and beyond. J Exp Bot 2011;62:3713–25.

[46] Dunoyer P, Melnyk C, Molnar A, Slotkin R. Plant mobile small RNAs. Cold Spring Harb Perspect Biol 2013;5. 017897-XXX.

[47] Fujimoto R, Uezono K, Ishikura S, Osabe K, Peacock WJ, Dennis ES. Recent research on the mechanism of heterosis is important for crop and vegetable breeding systems. Breed Sci 2018;68(2):145–58. https://doi.org/10.1270/jsbbs.17155.

[48] Greaves IK, Gonzalez-Bayon R, Wang L, Zhu A, Liu P-C, Groszmann M, Dennis ES. Epigenetic changes in hybrids. Plant Physiol 2015;168(4):1197–205. https://doi.org/10.1104/pp.15.00231.

[49] Chen ZJ. Genomic and epigenetic insights into the molecular bases of heterosis. Nat Rev Genet 2013;14(7):471–82. https://doi.org/10.1038/nrg3503.

[50] Johannes F, Porcher E, Teixeira FK, Saliba-Colombani V, Simon M, Agier N, Colot V. Assessing the impact of transgenerational epigenetic variation on complex traits. PLoS Genet 2009;5(6):XXX–YZ. https://doi.org/10.1371/journal.pgen.1000530.

[51] Reinders J, Wulff BBH, Mirouze M, Marí-Ordóñez A, Dapp M, Rozhon W, Paszkowski J. Compromised stability of DNA methylation and transposon immobilization in mosaic *Arabidopsis* epigenomes. Genes Dev 2009;23(8):939–50. https://doi.org/10.1101/gad.524609.

[52] Teixeira F, Heredia F, Sarazin A, Roudier F, Boccara M, Ciaudo C, Cruaud C, Poulain J, Berdasco M, Fraga M, et al. A role for RNAi in the selective correction of DNA methylation defects. Science 2009;323:1600–4.

[53] Cortijo S, Wardenaar R, Colomé-Tatché M, Gilly A, Etcheverry M, Labadie K, Caillieux E, Hospital F, Aury JM, Wincker P, et al. Mapping the epigenetic basis of complex traits. Science 2014;343:1145–8.

[54] Crow JF. 90 years ago: the beginning of hybrid maize. Genetics 1998;148:923–8.

[55] Duvick DN. Biotechnology in the 1930s: the development of hybrid maize. Nat Rev Genet 2001;2:69–74.

[56] Cheng SH, Zhuang JY, Fan YY, Du JH, Cao LY. Progress in research and development on hybrid rice: a super-domesticate in China. Ann Bot 2007;100:959–66.

[57] Jackson S, Chen Z. Genomic and expression plasticity of polyploidy. Curr Opin Plant Biol 2010;13:153–9.

[58] Lauss K, Wardenaar R, Oka R, van Hulten MHA, Guryev V, Keurentjes JJB, Johannes F. Parental DNA methylation states are associated with Heterosis in epigenetic hybrids. Plant Physiol 2018;176 (2):1627–45. https://doi.org/10.1104/pp.17.01054.

CHAPTER 3

Understanding immune system development: An epigenetic perspective

Ayush Madhok, Anjali deSouza, Sanjeev Galande
Laboratory of Chromatin Biology and Epigenetics, Department of Biology, Indian Institute of Science Education and Research, Pune, India

Contents

Introduction: An outline of the chapter 39
Epigenetic modifications and their functional output 40
  Transcriptional control of epigenetic modifications: From on/off to cycling of epigenetic modifications 41
  DNA methylation: The mark of silence 41
  Histone modifications: Key epigenetic drivers 42
  Interdependence of DNA and histone modifications 43
  Epigenetic modifications mediated by chromatin remodeling and noncoding RNA (ncRNA) 44
  Accessible chromatin: A prelude to transcription 44
  Poised state of a gene: Epigenetic modifications priming gene for activation 44
Epigenetic mechanisms regulating immune cell development 45
  Epigenetic regulation during hematopoiesis 45
  Development of innate immune cells: Regulation by epigenetic processes 45
  Epigenetic mechanisms in adaptive immune system 51
Concluding remarks and future perspective 63
Acknowledgments 63
References 64
Further reading 76

Introduction: An outline of the chapter

About every cell of an organism harbors an identical genome, yet there is a great diversity in the gene expression profiles of various individual cell types. This diversification occurs during their development and is maintained stably thereafter. Epigenetic changes by definition do not alter the genomic sequence, but play a profound role in gene expression and DNA accessibility, via modulation of the chromatin state. Covalent modifications such as histone modifications or DNA methylation as well as noncoding RNA mediated epigenetic mechanisms that alter the chromatin state have been described. Most of these modifications are reversible [1].

Epigenetics of the Immune System
https://doi.org/10.1016/B978-0-12-817964-2.00003-4
© 2020 Elsevier Inc.
All rights reserved.

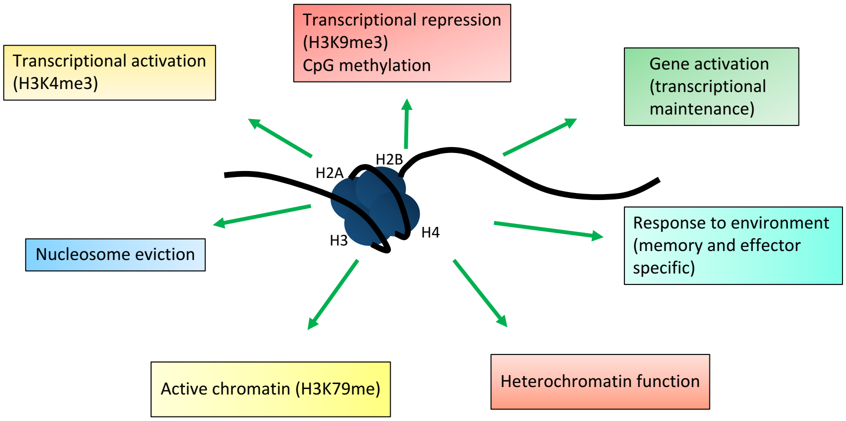

Transcriptional activation (H3K4me3)
Transcriptional repression (H3K9me3) CpG methylation
Gene activation (transcriptional maintenance)
H2A
H2B
H3
H4
Nucleosome eviction
Response to environment (memory and effector specific)
Active chromatin (H3K79me)
Heterochromatin function

Fig. 1 Role of epigenetic mechanisms in transcriptional regulation. Changes to the "chromatin state" are spatiotemporally controlled by epigenetic modifications. These modifications not only start or stop gene expression, but also relay changes from the environment to gene activation/repression; for example, activation of naïve T cells via antigen exposure leading to diverse gene expression profiles.

Epigenetic regulation of transcription is dynamic, i.e., it does not follow the binary rule of transcription on/off state. Instead, it acts as a rheostat of expression kinetics, of maintenance, and of the state of poised gene expression. Epigenetic modifications can also be influenced by exogenous cues, thereby adding another dimension to epigenetic control of transcription. Different stages of immune cell development including early progenitor commitment, lineage fate decision and selection, differentiation, as well as effector and memory functions are regulated by epigenetic changes (stage-specific roles).

This chapter begins with an outline of the basic epigenetic mechanisms, resulting in key modifications, and their physiological effects (Fig. 1). Further, it describes the role of epigenetic mechanisms in immune cell development using specific cell types from both innate and adaptive immunity. Subsequent subsections will talk about transcriptional control and development as well as functional aspects of immune subsets that are driven by epigenetic changes.

Epigenetic modifications and their functional output

A eukaryotic genome is highly compacted, and the primary unit of packaging of DNA is a highly organized entity called the nucleosome. Each nucleosome consists of an assembled octamer of histone proteins H2A, H2B, H3, and H4 around which 146 base pairs of DNA are wound [2]. There are many known epigenetic modifications associated with these histones, occurring mainly on their free N-terminal tails such as methylation, acetylation, ubiquitination, sumoylation, citrullination, and phosphorylation—collectively termed as posttranslational modifications (PTMs) (Fig. 1). In contrast, the nucleosomal

DNA is only associated with one modification in the form of DNA methylation or its chemical variations. Other mechanisms of epigenetic changes involve a noncoding RNA (micro, small interfering, and long noncoding RNA) mediated transcriptional switch [3]. As stated above, epigenetic changes act in a "nonbinary" fashion. In the section below, we discuss how DNA modifications as well as key PTMs of histones (majorly those of H3 and H4) coordinately regulate the transcriptional status of a cell.

Transcriptional control of epigenetic modifications: From on/off to cycling of epigenetic modifications

There is temporal and spatial expression of different genes in different cell types in spite of the genome being identical in all cell types of an organism. This fundamental aspect of cell identity requires timely accessibility of nucleosomal DNA which otherwise is largely inaccessible by a majority of transcription factors. Various changes in the higher-order chromatin structure precedes transcriptional activation such as nucleosomal repositioning, removal of repressive histone modifications, and deposition of activating histone modifications. To execute these functions, many different epigenetic readers, writers, and erasers constantly cycle on various genomic loci [1], as discussed in "Histone modifications: Key epigenetic drivers" section.

DNA methylation: The mark of silence

The most widely known epigenetic mark to modify DNA is the covalent attachment of a methyl group to the Carbon fifth position of cytosine almost exclusively at CpG dinucleotides in mammals. DNA methylation is generally considered transcriptionally repressive and is especially observed at the CpG dinucleotides concentrated at the promoter regions (CpG islands) and present in other regulatory regions [4–6]. Primarily, DNA methylation acts as a block for transcription factors to recognize and bind the DNA, thus rendering them inactive [7].

DNA methylation is carried out by various enzymes with methyltransferase activity such as DNA methyltransferase 1 (DNMT1), DNMT3A, and DNMT3B. DNMT1 recognizes hemi-methylated DNA (i.e., recognition of methyl groups on the parent DNA strand) and adds methyl residues to the newly replicated daughter DNA strand, thereby maintaining the methylation status in the daughter strand as well (Fig. 2A). Conversely, DNMT3A and 3B methylates DNA de novo, i.e., methylation of naked DNA [3]. DNA methylation is not a permanent epigenetic mark and can be "actively" reversed. Removal of the methyl group is mediated by a family of well-characterized enzymes called ten eleven translocation 1 (TET1), TET2, and TET3. They carry out a multistep process: oxidative conversion of 5-methylcytosine to 5-hydroxymethylcytosine, subsequent formation of 5-formyl or 5-carboxycytosines, and finally action by DNA glycosylases to

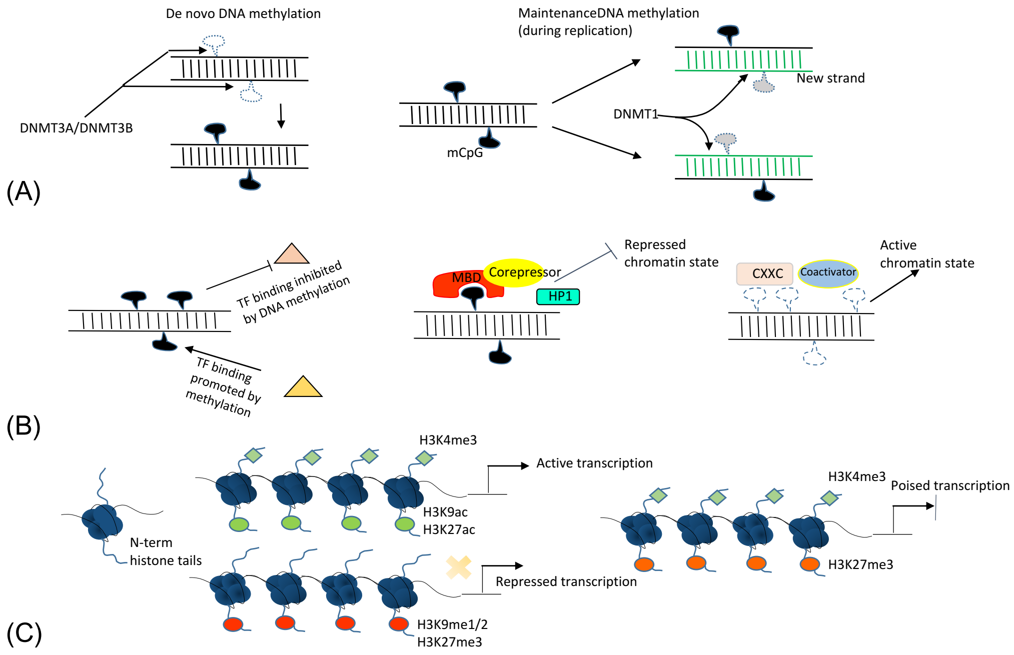

De novo DNA methylation
DNMT3A/DNMT3B
MaintenanceDNA methylation
(during replication)
New strand
DNMT1
mCpG
(A)
TF binding inhibited
by DNA methylation
TF binding
promoted by
methylation
MBD
Corepressor
HP1
Repressed
chromatin state
CXXC
Coactivator
Active
chromatin state
(B)
H3K4me3
Active transcription
H3K9ac
H3K27ac
N-term
histone tails
Repressed transcription
H3K9me1/2
H3K27me3
H3K4me3
Poised transcription
H3K27me3
(C)

Fig. 2 Various epigenetic mechanisms drive gene expression. (A) DNA methylation, generally, is repressive to gene expression. DNMT3A and DNMT3B complexed with DNMT3L drives de novo methylation, which is maintained during replication by DNMT1 methyltransferase. (B) Methylation (black spheres) in the CpG DNA context can recruit certain methyl-binding domain (MBD)-containing proteins, which further recruit chromatin modifiers to suppress transcription or mediate heterochromatinization (via HP1). Furthermore, CpG islands at promoters of active genes are protected by CFP1 (CXXC) from getting methylated. (C) Histone modifications which switch "on/off" transcription are depicted. H3K4me3 and H3K27ac, and H3K27me3 and H3K9me are generally associated with active and inactive chromatin, respectively. On certain gene loci, the histones contain a dual mark of H3K4me3 and H3K27me3 which allows of a "poised" state, and is swiftly activated (gene) upon external stimulus.

excise the additional bases, completing the process of demethylation [8, 9]. Detailed mechanisms of DNA methylation and demethylation are covered in other chapters.

Histone modifications: Key epigenetic drivers

Histones allow for DNA compaction and are also highly modified to mediate stage dependent or external cue-responsive transcription. Histone acetylation is correlated with gene activity. Histone Acetyltransferases (HATs) and coactivators such as PCAF (P300/CBP-associated factor) and p300 are responsible for acetylation of histones (H3 and H4) on their positively charged N-terminal lysine residues, thereby neutralizing positive lysines. This in turn reduces the positive charge of histones resulting in loosening of DNA on the core histone proteins [10, 11]. Consequently, the DNA is rendered available for transcription factor binding. Like DNA methylation, the posttranslational modifications of histones are reversible. Histone acetylation is removed catalytically by another family of enzymes called histone deacetylases or HDACs. There is a dynamic cycling of HATs and HDACs mediating a transcription on/off state resulting in temporal control of gene expression [1]. Histone acetylation, which signifies an active gene state, has been extensively studied on lysine 27 or lysine 9 of H3 [12].

Another important modification is methylation at key lysine residues of H3 and H4 histones. Methylation, as opposed to acetylation, may lead to either activation or repression depending on which lysine residue is modified (Fig. 2). Methylation at H3K4, especially at the promoter region, as well as at H3K79 and at H3K36 in the gene body, is transcriptionally permissive [1, 13]. Methyltransferases of family MLL and Set 1/7/9 are responsible for H3K4 methylation, whereas lysine methyltransferase (KMT) DOT1L and NSD1 methylate H3K79 and H3K36, respectively [3, 13, 14]. In contrast, methylation at H3K27 and H3K9 as well as at H4K20 is transcriptionally repressive. Di or tri-methylation at H3K27 is carried out by polycomb repression complex (PRC2) protein enhancer of zeste 2 (EZH2) [15–17]. Lysine methyltransferases SUV39H1/H2 and G9a carry out H3K9 methylation (both di and tri) [18–20] and SUV4-20H1 mediates H4K20 methylation [21]. Increasing evidences indicate that the histone methylations discussed above are reversible [22] via the action of lysine demethylases. For example, Jumonji (JmjC) domain protein JMJD2A is responsible for H3K9 di or tri-methylation [3, 13].

Interdependence of DNA and histone modifications

DNA methylation and histone modifications have many nexus points that together co-regulate the chromatin state. The DNA binding protein MeCP2 and other members [23] specifically recognize methylated CpGs and recruit HDACs, which further the repressive state of the gene via histone deacetylation. Using embryonic stem cells, it has been shown that a complex of HDACs and the methyltransferase G9a are involved in first deacetylation of the lysine residue of Histone H3 (H3K9), followed by methylation of H3K9 [24, 25]. Furthermore, the di/tri-methylated H3K9 is specifically recognized and bound by HP1, a chromodomain containing protein, resulting in local heterochromatinization [25, 26] (Fig. 2B). Methyltransferase G9a also recruits the enzymes DNMT3A and DNMT3B that catalyze de novo methylation of the promoter DNA [26].

DNA methylation and histone modifications can also be anticorrelated to each other. This is observed at CpG islands of active genes which remain unmethylated [25], primarily due to sequence specific binding of CXXC-type zinc finger protein 1 (CFP1) [27, 28]. Binding of CFP1 allows recruitment of lysine methyltransferases (KMTs) that methylate H3K4. Mono, di, or tri-methylation of H3K4 inhibits binding of DNMT3L to the DNA which would otherwise recruit methyltransferases DNMT3A/3B onto DNA. Thus, de novo methylation at CpG islands of active genes is prevented [29].

Epigenetic modifications mediated by chromatin remodeling and noncoding RNA (ncRNA)

Nucleosome remodeling, which includes histone sliding, removal, or incorporation, is an ATP-dependent process involving various chromatin-remodeling proteins [30, 31], resulting in better accessibility of DNA to specific transcription factors. Chromatin access, which involves nucleosomal sliding and nucleosome eviction or ejection, is mainly carried out by SWItch Sucrose nonfermentable (SWI/SNF) complex remodelers. Nucleosome editing is carried out by INO80 subfamily proteins. These selectively remove particular histones from nucleosomes and replace them with either canonical or variant histones. The addition or removal of particular histone residues affects the recruitment of transcription factors and their activity [32].

Noncoding RNAs comprising both short (miRNAs, piRNAs, and siRNAs) and long noncoding RNAs (lncRNA) have also been implicated in driving epigenetic mechanisms and transcriptional regulation [33–35]. For instance, the lncRNA HOTAIR (HOx Transcript Antisense RNA), which is transcribed from the HOXC cluster, can silence a 40 Kb HOXD cluster located afar, via recruitment of the PRC2 complex, resulting in heterochromatinization and repressive H3K9 tri-methylation [36–38]. Similarly, another study showed that piRNA efficiently brings HP1a protein to particular genomic loci, leading to gene silencing [39].

Accessible chromatin: A prelude to transcription

DNA is sturdily associated with nucleosomes, making it inaccessible to activating modifications (described above). Therefore, certain gene-regulatory elements are devoid of nucleosomes, facilitating both the addition of DNA/histone modifications and an increased susceptibility to nuclease digestion. The accessibility of these elements is demonstrated by the DNase hypersensitivity assay followed by sequencing, and also by a fairly new technique called ATAC sequencing (Assay for Transposase-Accessible Chromatin sequencing) [40, 41]. These methods enable genome-wide mapping of accessible chromatin elements such as promoters, enhancers and can predict sites for transcription factor binding [40]. Studies have shown that chromatin-openness relates well with cell type-specific transcription factor binding events at proximal and distal regulatory elements [42]. Additionally, there exist "pioneer" transcription factors that can bind to closed or repressed chromatin (or nucleosome) and can gradually open the chromatin for activation [42].

Poised state of a gene: Epigenetic modifications priming gene for activation

A poised gene state is characterized by the presence of a "bivalent epigenetic mark" with both activating and repressive histone modifications H3K4me3 and H3K27me3, respectively, at the same genomic locus [43] (Fig. 2C). This "poised" chromatin state was first observed at promoters of developmental genes in embryonic stem (ES) cells [43]. Since then, many studies have corroborated the importance of "bivalent" modifications in regulating developmentally important genes in pluripotent progenitors and in germ cells [44]. At initial stages of development, a poised gene promoter enables the gene to respond immediately to extracellular stimuli, especially to various differentiation cues. As soon as the signal is received, active demethylation of H3K27 ensues, methylated H3K27 being repressive in nature. This leads to an instantaneous transcription initiation due to the residual H3K4 mark. Although in lower proportions, unipotent cells such as the T cells, MEFs, and other cell populations also exhibit "bivalency" with coexisting H3K4 and H3K27 tri-methylation marks. [45, 46]. Studies using advanced techniques such as ATAC seq and Hi-C have elucidated the dynamic modulation of chromatin in specific immune lineages during their development [40].

Epigenetic mechanisms regulating immune cell development

Immune cell development begins with the differentiation of a multipotent hematopoietic stem cell (HSC) toward the lymphoid or myeloid lineages, and is accompanied by a gradual decline in the developmental potential or "stemness" of the cell. These lymphoid and myeloid precursor cells pass through distinct stages of development till they are terminally differentiated into mature immune cells of innate or adaptive fate. These processes involve the cooperative action of lineage-specific transcription factors and epigenetic modifiers that together switch on key genes of a particular lineage, while simultaneously repressing genes that are active in precursor cells or in other immune cell lineages.

Epigenetic regulation during hematopoiesis

Multiple developmental processes such as stem cell renewal, cell fate decisions, and differentiation of effector and memory immune cells [47] are regulated by epigenetic mechanisms. In particular, the self-renewing capacity of hematopoietic stem cells is highly dependent on DNA methylation, which is correlated with a repressed gene state. This was demonstrated in DNMT1 knockout mice, wherein the absence of DNMT1 leads to premature differentiation skewed toward myeloid lineages [48]. In contrast, demethylation of DNA via TET proteins is important for timely activation of self-renewal factors. Mutations in TET family proteins result in skewed differentiation to myeloid subsets [49]. Similarly, mutations in the PRC1 complex proteins, which endow H3K27 methylation marks, result in the loss of self-renewal capacity of HSCs.

Development of innate immune cells: Regulation by epigenetic processes

Innate immune cells are widely known as the "first line of defense" against infection and mediate very strong responses at the initial stages of an infection. Innate immune cells are varied and include neutrophils, basophils, macrophages, dendritic cells, natural killer cells, and innate lymphoid cells (ILC1, ILC2, and ILC3 cells), among others. The complementary roles of lineage transcription factors and chromatin modifiers confer specific gene expression profiles for each cell type. This section delves into epigenetic changes and the action of various "lineage determining factors" in the development of ILCs, natural killer (NK) cells, dendritic cells, and macrophages.

Development of natural killer (NK) cells

Discovered more than four decades ago, NK cells are the founding members of the innate immune system and belong to the ILC1 subset [50, 51]. NK cells are critical for host defense against virus-infected cells and against tumor formation due to their rapid release of inflammatory cytokines, namely INFγ and TNF-α, and of toxic proteins such as granzyme B and perforins. Like B and T cells, NK cells also originate from common lymphoid progenitor cells (CLP) (Fig. 3). Similar to B cells, they develop primarily in the bone marrow [52]. NK cells are unique in that they can recognize foreign or transformed "self" cells [53] that have downregulated MHC-I. However, they also express inhibitory receptors called Killer-cell Immunoglobulin-like Receptors (KIR), which recognize MHC class-I expressed on self-cells and thus prevent a "self-attack" by NK cells [54]. While epigenetic modifications guiding development of NK cells are poorly characterized, the role of transcription factors contributing to NK cell development is more clearly

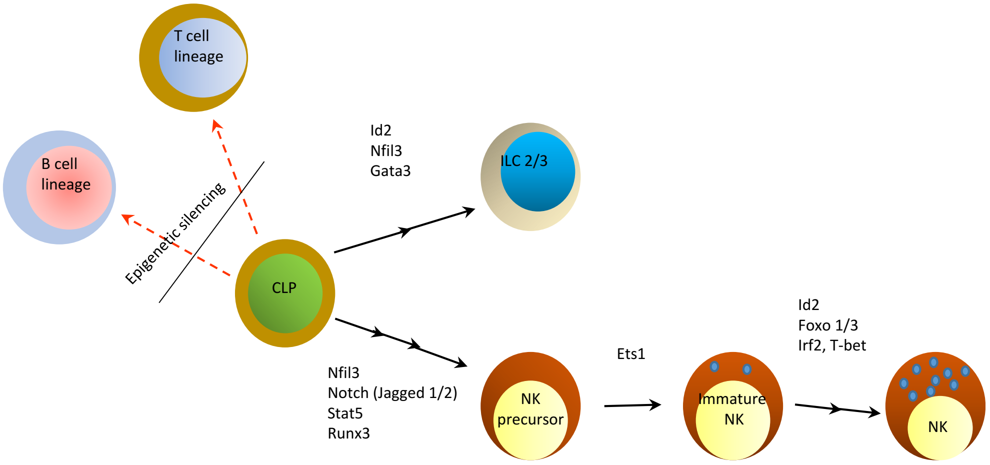


Fig. 3 Development of NK cells from common lymphoid progenitors (CLPs). NK cells are developed from the same progenitors as B or T cells. The epigenetic program of NK cell precursors initially shuts off the B- or T-cell program by repressing (via H3K27me3 modifications) their lineage specifying TFs *Pax5* and *Tcf7,* respectively. Additionally, expression of the ILC specific factors increases in which Jagged receptor mediated Notch signaling is important for NK cell development along with STAT5 activity. Upon lineage commitment, Ets1, T-bet, and Irf mediated maturation and effector phenotype of NK cells is shaped.

defined [52, 55, 56]. The epigenetic mechanisms presently known to mediate differentiation and function of NK cells are described below.

The initial studies exploring the epigenetic regulation in NK cells focused on genes encoding KIR inhibitory receptor, and showed that the promoters of KIR genes in hematopoietic progenitor cells (HPCs) and non-KIR expressing cells are heavily methylated at CpG residues, but are demethylated at these regions in NK cells and other KIR expressing cell types [57, 58]. In contrast, histone acetylation does not seem to play a role in KIR gene expression since culture of NK cells with an HDAC inhibitor such as valproic acid had no effect on KIR expression [59, 60], but there was reduced NK cell cytotoxicity observed with such treatment [59]. In addition, the expression of NKG2D (key activating receptor on NK cell surface) [61] has been shown to be regulated by epigenetic modifications of DNA and histones. Accordingly, DNA demethylation in conjunction with H3K27 demethylation and H3K9ac resulted in the activation of NKG2D, while incubation of NK cells with HATi led to the downregulation of NKG2D. Development of NK cells from CLPs is primarily driven by Notch signaling, specifically through Jagged1 and Jagged2 receptors [62, 63] and other important transcription factors such as Id2 and Nfil3 (E4BP4) [64–69]. In particular, Nfil3 is one of the most crucial TF for NK cell lineage commitment (Fig. 3), since Nfil3-deficient mice do not have NK cells, but are normal for B- and T-cell development [64, 70]. Furthermore, signaling downstream of the cytokine IL-15 is critical for Nfil3 expression in NK precursors [71, 72], mediated in part through STAT5 transcription factor. It was exemplified by deletion of Stat5b, which resulted in complete loss of NK cells [73, 74]. Other transcription factors like Runx3 and Cbfβ are also required for NK cell development in liver chimeras mediated through expression of Il2Rβ [75].

The effector function of NK cells is mediated by key effector molecules—granzyme B and perforins. Production of perforin is dependent on Ets4 (myeloid Elf1-like factor), which occupies and activates two promoters on the *Prf*1 gene, [76]. Also, TF STAT5 has been shown to bind directly on the upstream enhancers of *Prf*1 [75], thus indicating a role for STAT5 in both NK cell development, as described above, and in NK cell effector function. NK cells rapidly produce interferon-γ (INFγ) upon stimulation. Production of IFN-γ in NK cell precursors is dependent on consistent acetylation at a distal conserved noncoding sequence (CNS) in the IFN-γ locus, corresponding to a 22 Kb upstream enhancer region called CNS-22, which opens the locus via binding of t-bet transcription factor [77].

ILC2 cell development

Type 2 innate lymphoid cells or ILC2 cells mediate type 2 innate immune responses against helminth and viral infections and allergies [78, 79]. ILC2 cells express receptors for IL-25 and IL-33 and thymic stromal lymphopoietin (TSLP) cytokines, and upon stimulation secrete type 2 cytokines like IL-4, IL-5, IL-9, and IL-13 during infection [79–81].

Many transcription factors such as GATA3, RORα, TCF1, and GFI1 play important roles in the development of ILC2 cells [82–85] (Fig. 3). Among these, GATA3 is considered to be the key driver of ILC2 cell fate [83–85]. Transgenic overexpression of GATA3 led to an induction of IL-33 receptor expression, and this was concomitant with higher numbers of ILC2 cells [83]. TCF1 is also critically essential for ILC2 development [86, 87] since mice lacking TCF1 (TCF1 KO) do not develop ILC2 cells, resulting in an impaired innate immune response. Furthermore, GFI1 knockout mice (GFI1 KO) have impaired ILC2 function since GFI1 transactivates both IL-33 and IL-25, and also controls GATA3 expression [88]. Although the role of transcription factors in the development and function of ILC2 cells has been addressed, the epigenetic modulations associated with gene regulation in these cells is still underway.

Development of macrophages

Macrophages develop from hematopoietic stem cells and belong to the myeloid lineage of the immune system. They act in the innate arm of defense and specialize in the process of phagocytosis, clearing infected cells and debris [89]. Macrophages reside in almost all tissues and are associated with maintaining tissue homeostasis; upon infection, macrophages get activated and release inflammatory chemokines, resulting in microbe phagocytosis and local inflammation. Depending on the microenvironment, activated macrophages polarize to specialized subsets [90, 91]. Epigenetic changes have been ascribed to both development and polarization of macrophages [92] and are discussed in this section.

Early during HSC differentiation, the myeloid lineage segregates from lymphoid lineage and as cells develop toward macrophages, the epigenetic landscape specific for progenitor cells is reconfigured to that accompanying macrophage development. Similar to "bivalent promoters" described in ES cells (discussed in "Poised state of gene: Epigenetic modifications priming gene for activation" section), numerous macrophage lineage genes exist in a "poised" or "bivalent" state of histone marks. In fact, via comparative analysis, it has been shown that about 61% enhancers in the granulocyte macrophage progenitors are "poised" or "bivalent" [93]. An example is the gene encoding myeloid specific CCAAT enhancer binding protein (C/EBP) which exhibits both H3K4me3 and H3K27me3 chromatin marks in HSCs [94]. Upon differentiation to myeloid lineage, the *Cebpa* (gene encoding C/EBP) promoter loses the repressive H3K27 methylation mark, but retains the permissive H3K4me3 methylation mark due to which there is induction of C/EBP expression. C/EBP in turn modifies the chromatin landscape promoting further differentiation [89]. PU.1 is a "bifunctional" member of the Ets family transcription factors, and is quite potent in driving myeloid lineage differentiation by maintaining the H3K4me1 mark specific for macrophages, facilitating an active chromatin state [95, 96]. Additionally, the expression levels of PU.1 in myeloid progenitors are key to macrophage differentiation since higher PU.1 to CEBP ratio results in macrophage development, whereas the converse gives rise to neutrophils [96]. Other transcription factors such as the members of the STAT and GATA family also play important roles in modulating the epigenetic landscape in macrophages in response to specific stimuli, as discussed below [97, 98].

Tissue macrophages polarize to M1 or M2 subtypes depending on activation by specific factors in the microenvironment [90, 91] (Fig. 4A). Classical activation of macrophages by lipopolysaccharides (LPS), IFN-γ, and various toll-like receptor (TLRs) ligands result in the M1 subtype which drives Th1 responses. Alternative activation of macrophages mediated by cytokines such as IL-4, IL-13, and IL-10 polarizes macrophages to M2 cells which exuberate Th2 responses [99]. M1 macrophages are efficient at killing microbes. These macrophages when polarized effectively produce inflammatory cytokines such as tumor necrosis factor (TNF), IL-12, and IL-1, and chemokines, namely CXCL9 and CXCL10 [100, 101]. M2 macrophages, on the other hand, are responsible for maintaining tissue function under physiological conditions and stress response, and promote wound repair and healing by reducing inflammation [102].

Epigenetic mechanisms regulating polarization and effector function of macrophages comprise DNA methylation and histone PTMs along with nucleosome remodeling. The inflammatory activation of M1 macrophage in response to TLRs has been shown to be primarily regulated by epigenetic processes [103–107] through signaling via mitogen-activated protein kinases (MAPKs), NF-κB, and IRFs, inducing key cytokine synthesis including TNF, IL-1β, IL-6, and IL-12. An epigenetic landscape has arisen to explain gene regulation at these loci, wherein lineage factors PU.1 and C/EBP directly bind to and open the chromatin at their key regulatory regions [108, 109] (Fig. 4B). Intriguingly, the promoters of resting macrophages are in a permissive state, harboring H3K4me3 and K3K27ac [105]. When devoid of TLR signaling, these macrophages restrain inflammatory and cytokine gene expression via recruitment of the repressor Bcl-6 and other nuclear receptors to these gene loci. This in turn recruits various HDACs and demethylases at promoter regions, thereby deprecating the amount of positive histone PTMs [103, 110]. Furthermore, these gene loci are enriched in repressive epigenetic marks H3K27me3 [104, 111], K3K9me3 [112, 113], and H4K20me3 [114].

The "repressive code" is erased upon TLR signaling via recruitment of demethylases such as JMJD3, JMJD2d, and AOF1 [104, 114], and ATP-dependent nucleosome remodeling by SWI/SNF complex [107, 115] (Fig. 4A).

M2 macrophage activation by helminthes infection (in vivo) is facilitated by JMJD3 (Jumonji domain containing-3, a histone demethylase), which reverses the inhibitory H3K27 tri-methylation modification at the *IRF4* locus [116]. Additionally, HDAC3 has been identified as a negative regulator of M2 macrophage activation through deacetylation of putative enhancers on IL-4-induced M2-specific genes [117].

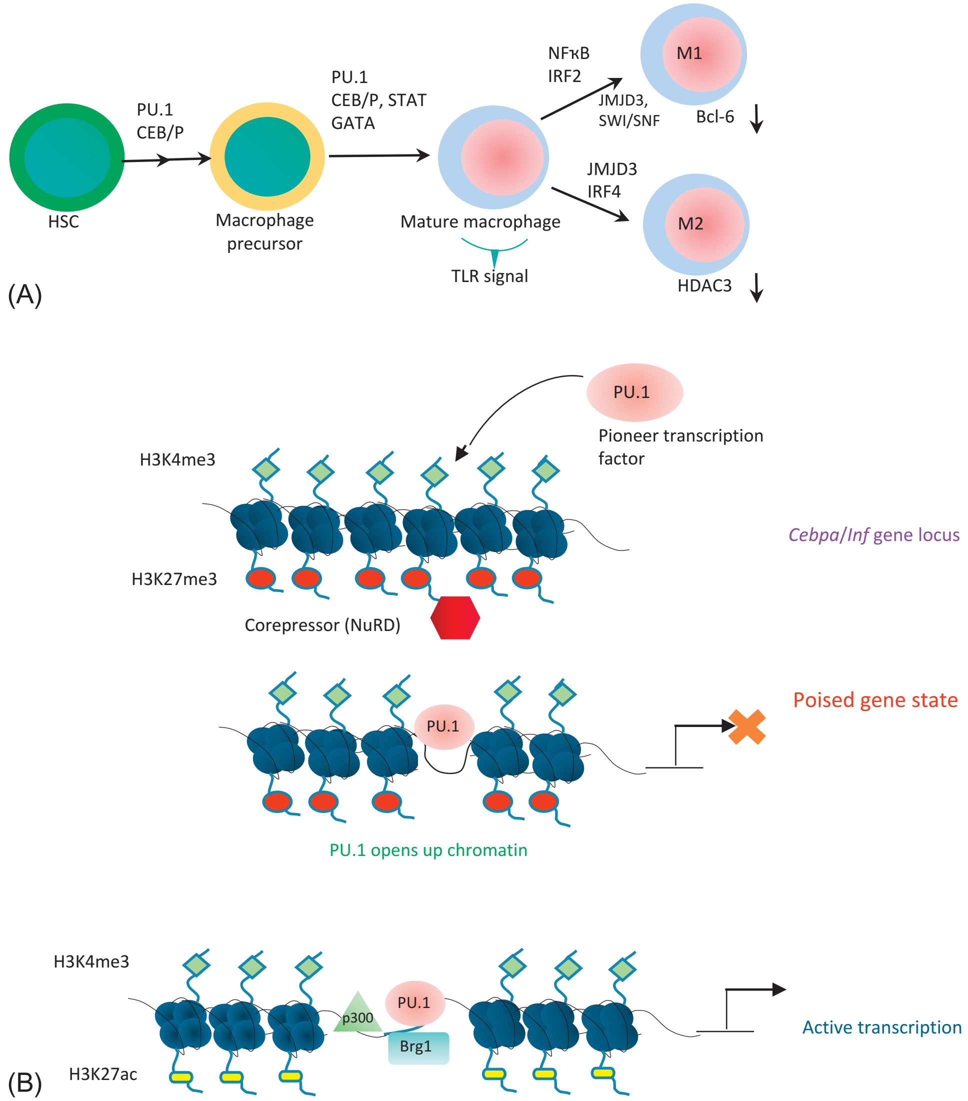

PU.1
CEB/P
HSC
Macrophage precursor
PU.1
CEB/P, STAT
GATA
Mature macrophage
TLR signal
NFκB
IRF2
JMJD3,
SWI/SNF
M1
Bcl-6 ↓
JMJD3
IRF4
M2
HDAC3 ↓
(A)
PU.1
Pioneer transcription factor
H3K4me3
H3K27me3
Corepressor (NuRD)
Cebpa/*Inf* gene locus
PU.1
Poised gene state
PU.1 opens up chromatin
H3K4me3
p300
PU.1
Brg1
Active transcription
H3K27ac
(B)

Fig. 4 Lineage commitment of macrophage from HSCs and epigenetic changes. (A) A schematic of macrophage development and polarization is depicted. Various transcription factors and epigenetic modifiers are involved in both the processes. Upon receiving toll-like receptor (TLR) signaling, mature macrophages get activated and polarize toward M1 or M2 phenotype. (B) The pioneer TF PU.1 is important in defining macrophage lineage from HSCs. On *Cebpa* (encoding CEB/P) as well as various cytokines' loci, PU.1 binds closed-chromatin and opens it up for recruitment of various coactivators such as HAT p300 and Brg1. Upon TLR signal, the repressive methylation marks are actively removed by TETs, rendering the genes active for macrophage response.

Development of dendritic cells

Dendritic cells (DCs) function as "professional" antigen-presenting cells to T cells, thereby initiating an adaptive immune response against a variety of pathogens and tumors [118, 119]. Under a physiological steady state, DCs are primarily categorized into two distinct populations, namely conventional DCs (cDCs) and plasmacytoid DCs (pDCs), based on their surface markers, location, and function [120]. During inflammation, DCs undergo dramatic changes in their phenotypes and certain subsets of DCs secrete cytokines such as IL-13 and IL-12 as well as Notch ligands like DLL4, which along with DC-T cell interactions induce various T-cell subtypes such as Th2, Th1, Th17, and T cytotoxic [121, 122].

Transcription factors and epigenetic mechanisms involved in DC development

Dendritic cells develop from HSCs through successive steps of lineage commitment: Hematopoietic stem cell to multipotent precursor (MPP) to common DC progenitor (CDP) to conventional DC or plasmacytoid DC [120, 123]. Additionally, DCs also differentiate from monocytes (alongside macrophages, as discussed above), called monocyte-derived DCs (moDC) [124]. Development of DCs from HSCs is associated with repression of genes such as GATA2, GFI1, and CCAAT/enhancer binding protein alpha (CBPa) which potentiate HSCs to nonhematopoietic lineages [125], and with upregulation of CDP specific genes such as HOXA1 and E2F2 [126]. Furthermore, both pDCs and cDCs have active expression of FLT3, STAT1, and IRF5, which are inhibited in MPPs, and are shown to mediate development and function of DCs. Among many TFs studied, PU.1 and STAT3 are shown to be the key lineage drivers of DCs. STAT3, for example, is involved in generation of both pDCs and cDCs from their precursors [127, 128], and deletion of STAT3 or targeted ablation of its ligand FLT3 results in impaired DC development in vivo [129]. PU.1, which is crucial for macrophage development as discussed previously, is also critical for DC fate. HSCs devoid of PU.1 have poor DC development potential [130].

The epigenetic landscape of DCs is formed prior to inflammation or pathogenic attack [131]. Both pDCs and moDCs are associated with H3K4me1 and H3K27ac marks associated with primed and active enhancers, respectively [126]. In the hierarchy of TF network from HSC to DC, PU.1 and CEBPB are pioneer TFs at several regulatory sites directing DC lineage commitment [126, 132]. Interestingly, ChIP-seq analysis of PU.1 and its colocalization with H3K4me1 showed an increase from 20% in MPPs to 70% in cDCs, demonstrating its ability to direct DC fate [126].

Epigenetic mechanisms in adaptive immune system

Adaptive immunity, unlike the innate counterpart, shows high specificity for an antigen and comprises mainly the T and B cells which express on their surface a unique cell surface receptor. Upon recognition of peptide antigen complexed with MHC, the receptor is ligated, resulting in clonal expansion of the cell, generating either plasma cell (antibody producing) in case of B cells, or helper T cell in case of T cells. Another notable feature of adaptive immunity is the generation of memory against a previously encountered pathogen. Thus, a secondary encounter with the pathogen results in a more rapid and stronger response compared to the primary encounter. The following sections describe the contribution of epigenetic changes to the development, differentiation, effector function, and memory formation of B and T cells.

B-cell development and differentiation—An epigenetic perspective

B cells are primary components of humoral immunity and are involved in antibody generation through plasma cells, cytokine production, and antigen presentation to T cells [133, 134]. The development of B cells begins from CLPs in the bone marrow independent of antigen, and is characterized by expression of cell fate determining factors and rearrangement of immunoglobulin chains—heavy (H) and light (L). CLPs progress through distinct stages of development, namely progenitor-B (pro-B), precursor-B (pre-B), and immature B cells [135]. The immature cells migrate out of the bone marrow to encounter the antigen, and undergo maturation, proliferation, and Ig class-switching. A portion of these mature B cells differentiate into memory cells which are long-lived and can mount a quicker and more efficient response on further antigen encounters. Many epigenetic mechanisms have been proposed for B-cell development, maturation, and effector function. Some of these are discussed in the following sections.

The early stages of B-cell development are regulated by stromal cytokines in the bone marrow, and only a limited set of transcription factors such as E2A, EBF, and PAX5 promote their differentiation from CLPs to pre-pro-B cells [136, 137] (Fig. 5A). E2A and EBF1 cooperatively activate B cell-specific genes in pre-pro-B cells [138, 139] and thereafter PAX5 acts as a B-cell lineage commitment factor restricting the potential of these cells to B-cell fate [140]. Transition from CLPs to pro-B cells requires the activity of B cell-specific cMYB transcription factor (encoded by *mb-1*). The promoter of *mb-1* is initially methylated at CpGs and is progressively demethylated by the action of E2A and EBF1 [141], allowing PAX5 to bind and activate the gene. Additionally, the chromatin remodelers SWI/SNF and NuRD complexes were shown to have positive and negative roles in mb-1 transcription, respectively, by altering chromatin accessibility [142].

A number of studies have elucidated the role of PAX5 in B-cell lineage commitment by repressing genes of other lymphoid lineages and activating B-cell-specific genes [143]. Accordingly, transcriptionally active histone marks such as H3K9ac, H3K4me2, and H3K4me3 are enriched on many of PAX5 direct targets [144–146]. PAX5 maintains an active gene status by interacting with chromatin remodeler BAF complex, with a HAT called CBP, and with PTIP protein which recruits H3K4 methyltransferase complex (MLL family proteins) [145]. Conversely, PAX5 also recruits NcoR1 corepressor

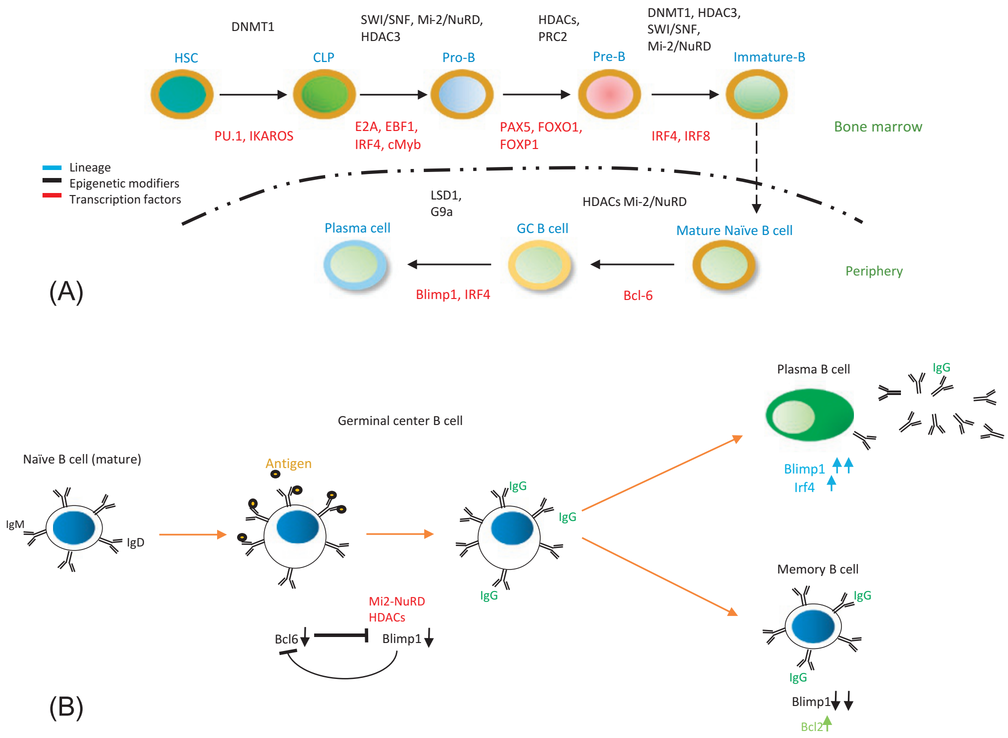

DNMT1
SWI/SNF, Mi-2/NuRD, HDAC3
HDACs, PRC2
DNMT1, HDAC3, SWI/SNF, Mi-2/NuRD
HSC
CLP
Pro-B
Pre-B
Immature-B
PU.1, IKAROS
E2A, EBF1, IRF4, cMyb
PAX5, FOXO1, FOXP1
IRF4, IRF8
Bone marrow
Lineage
Epigenetic modifiers
Transcription factors
LSD1, G9a
HDACs Mi-2/NuRD
Plasma cell
GC B cell
Mature Naïve B cell
Periphery
(A)
Blimp1, IRF4
Bcl-6
Plasma B cell
IgG
Germinal center B cell
Naïve B cell (mature)
Antigen
IgM
IgD
IgG
IgG
IgG
Blimp1
Irf4
Mi2-NuRD
HDACs
Bcl6
Blimp1
Memory B cell
IgG
IgG
Blimp1
Bcl2
(B)

Fig. 5 Transcriptional and epigenetic regulation of B-cell development and activation. (A) Stages of B-cell commitment are shown, with key TFs important for each stage of development. Various epigenetic modifiers also function in a stage-specific manner, the deficiency of which results in poor lineage commitment. (B) When a mature B cell (naïve) encounters an antigen in the germinal center (GC), it undergoes various epigenetic changes to facilitate the surface expression of antigen-specific immunoglobulins (IgGs). Bcl-6 is a critical regulator of B-cell activation, which suppresses Blimp-1 before activation via recruitment of Mi2-NuRD complex modifiers. In plasma B cells, Blimp-1 is the key factor along with Irf4 to expression of various plasma cell-specific gene expression and function. Memory cells, on the other hand, have suppressed expression of Blimp-1, with a concomitant increase in Bcl-2 expression—allowing for their long-lived phenotype and response function on antigen reencounter.

complex and the associated HDAC3 to genes like Ccr2 and Ncf4, leading to their repression [146] and allowing for unperturbed B-cell commitment.

Another important transcription factor required for B-cell development is Foxo1. Deletion of Foxo1 in mouse results in a developmental block at the pro-B stage [147]. Foxo1 is a "pioneer transcription factor" binding to its target genes while the chromatin is still in a condensed form, resulting in decondensation of chromatin and transactivation of genes [148]. Another study demonstrated the importance of "pioneer TF" Foxo1 in transition from pre-pro-B to pro-B cells. The pioneer TF is activated only when the initiator factor E2A gains the enhancer mark H3K4me1, which invokes the deposition of activation mark H3K4me3 on the promoters of B-cell lineage target genes [139].

Pro-B to pre-B cell commitment

MicroRNAs have been shown to play a vital role in early B-cell development; for example, miR-150 is expressed at low levels in HSCs, but is expressed at higher levels in B- and T-cell precursors [149]. Deletion of miR-150 in B-cell progenitors results in dysregulation of c-Myb (encoded by *mb-1*, discussed above), leading to impaired B-cell development and responses [150]. Further, Dicer, which is the key enzyme for mi- and si-RNA generation, is also implicated in early B-cell development. Conditional deletion of Dcr1 (encoding DICER) in the earliest B-cell precursors leads to a developmental block during pro-B cell stage-switch (pro- to pre-B) [151].

Pre-B to immature B cells—Formation of BCR

The B-cell receptor (BCR) is expressed on the B-cell surface and is formed by highly ordered V(D)J recombination events in loci encoding the receptor. Recombination between V, D, and J gene segments produces a vast array of "clonally distinct" receptors, recognizing diverse antigens. Generation of BCR is regulated by epigenetic modifications including changes in chromatin accessibility and modifications to DNA and H3 histones such as the permissive histone marks of H3K4me2, H3K4me3, H4Ac, and H3K79me2 [152–157]. These modifications help in the recruitment and binding of different modifiers at the Ig loci. The recombination activating gene (RAG) 2, via its PHD domain, recognizes H3K4me3 at the loci to be recombined via VDJ recombination [158]. DNA methylation also drives the status of VDJ recombination in pre-B cells. It has been observed that methylation inside the heptameric recombination signal recognized by RAG proteins markedly diminishes V(D)J recombination even if the RAG (1/2)-DNA complex is unperturbed [159].

Peripheral differentiation of B cells

Peripheral differentiation of B cells begins when a B cell expressing a fully functional BCR on its surface recognizes a specific antigen in the periphery. Germinal center (GC) is the peripheral niche in which B cells proliferate, switch Ig class, and undergo somatic hypermutation. A portion of these GC B cells that recognize antigen with high affinity are differentiated into long-lived memory or plasma B cells [160, 161]. Plasma cells continually secrete an antigen-specific antibody (previously, the BCR), whereas memory cells efficiently initiate quicker secondary immune responses mediating faster clearance of the pathogen. These differentiation stages are primarily regulated by two mutually exclusive repressors BCL6 and B lymphocyte induced maturation protein-1 (BLIMP-1) (Fig. 5B). Upon antigen encounter, B cells that do not express Bcl-6 undergo differentiation into plasma cells [162, 163]. The repressive function of Bcl-6 is mainly facilitated by the recruitment of Mi-2/NuRD chromatin remodeling complex and accompanying HDAC1/HDAC2 to many plasma cell-specific genes, preventing these cells from differentiating into plasma cells [164, 165] (Fig. 5B). The gene encoding the key plasma B-cell factor Prdm1 is also silenced via Bcl-6 bound HDACs [166]. Blimp-1, on the other hand, is shown to interact with demethylase LSD1, which correlates with histone modifications for open chromatin, thus facilitating the expression of plasma B cell-specific genes [167]. Furthermore, Blimp-1 acts as a transcriptional repressor of Bcl6, Pax-5, and Spib, by deacetylating promoters of these genes via activity of associated HDACs, thereby ensuring differentiation into plasma cells and inhibition of other cell fates [168, 169].

Role of epigenetic regulation in T cells

T cells constitute the cell-mediated arm of adaptive immunity, and are characterized by the membrane expression of either a αβ or γδ T cell receptor (TCR). Upon differentiation, αβ T cells give rise to cytotoxic $CD8^+$ T cells, helper $CD4^+$ T cells, and $Foxp3^+$ regulatory T cells (Tregs). CD8 T cells are cytotoxic killers that efficiently attack infected cells. CD4 cells, on the other hand, are "helpers," primarily of Th1, Th2, and Th17 subtypes. Treg cells keep the immune system under control, preventing an uncontrolled immune response against pathogens and preventing autoimmune responses against self-cells. A number of studies have shown that T-cell development, differentiation, and function are regulated by an intricate meshwork of epigenetic modifications and transcription dynamics.

From CLPs to T cells—Regulation by transcription factors and epigenetic modifications

T-cell development commences with the migration of common lymphoid progenitors (CLPs) from the bone marrow (BM) to the thymus—a distinct compartment for T cell selection and maturation. The CLPs are initially at the double negative (DN) stage and have the potential to develop into T lymphocyte, B lymphocyte, or myeloid lineages [170, 171]. In the thymic microenvironment, these progenitors respond to Notch signaling, usually at the DN2a and DN2b stages, and get committed to the T-cell fate [172–174] (Fig. 6). At the DN stage, rearrangement of the T cell receptor (TCR) beta chain occurs, followed by a critical checkpoint of beta selection wherein cells with wasteful TCR beta rearrangement are prevented from developing to the double positive (DP) stage [175, 176]. The thymocytes with productively rearranged TCRβ now mature to the DP stage, with commencement of the TCRα chain rearrangement. The T cell with a fully rearranged TCR dimer (αβ) undergoes both positive and negative selection [177–179]. DP thymocytes having a moderate affinity interaction with self-MHC-peptide complexes are positively selected and can further develop to single positive

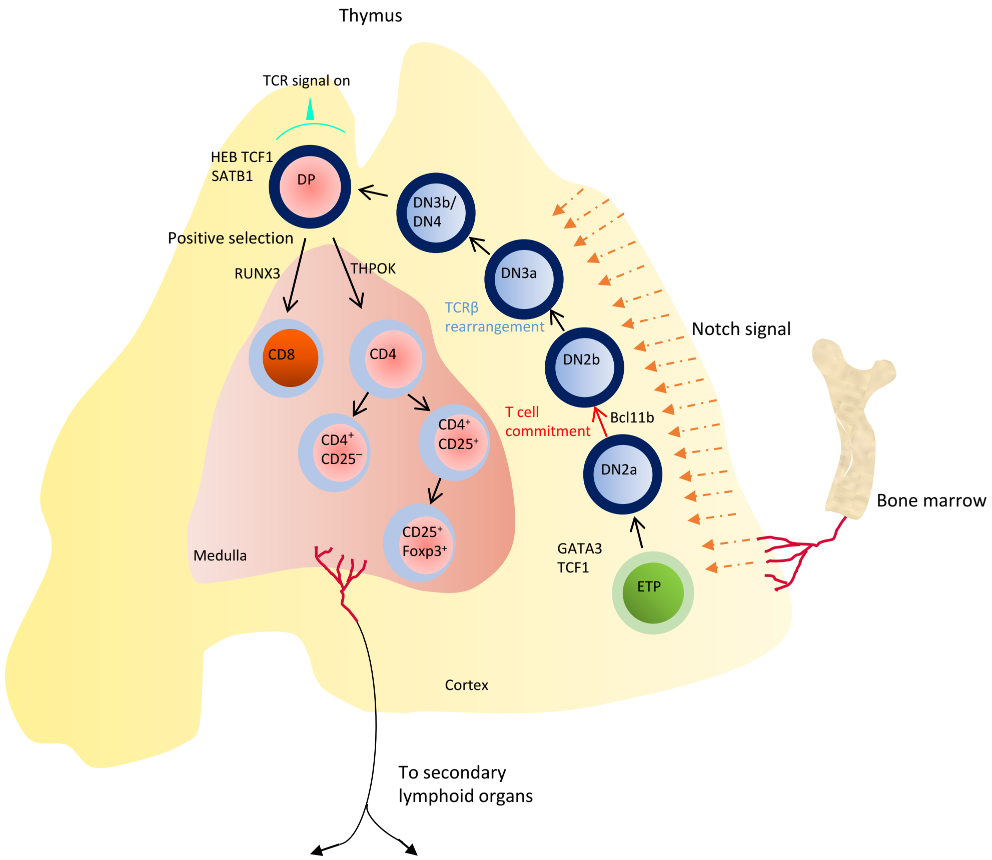

Thymus
TCR signal on
HEB TCF1
SATB1
DP
DN3b/
DN4
Positive selection
RUNX3
THPOK
DN3a
TCRβ
rearrangement
Notch signal
CD8
CD4
DN2b
T cell
commitment
Bcl11b
CD4+
CD25−
CD4+
CD25+
DN2a
Bone marrow
CD25+
Foxp3+
Medulla
GATA3
TCF1
ETP
Cortex
To secondary
lymphoid organs

Fig. 6 Transcriptional control of thymic T-cell development. Schematic representation of various stages of T-cell development from the emigrating CLPs from the bone marrow. Until the DN2a stage, the early thymic progenitors (ETPs) have the potential to give rise to other lymphoid subtypes (such as NK cells). Commitment to T-cell fate is achieved only at the DN2b stage via the activity of Bcl11b TF, along with TCF1 and GATA3. The continuous exposure to thymic Notch (DLL) signaling to DN cells ensures only T-cell fate. At the DP stage where the TCR is fully rearranged, various TFs cooperate to mediate differentiation into CD4 or CD8 stages. TCF1, HEB, and SATB1 are prominent TFs and chromatin regulators at the DP stage. SATB1 in particular is involved in chromatin looping into distinct transcriptional networks, allowing for TCR mediated gene expression and CD4 differentiation. Some CD4 cells are then based on the strength of the TCR signal, differentially express Foxp3, and develop into regulatory T cells. Mature naïve T cells (which are TCR responsive and functional) migrate to peripheral (secondary) lymphoid tissues (such as the spleen and lymph nodes), where they execute their effector or regulatory functions.

(SP) cells. Positive selection acts as a survival signal to these thymocytes [179]. DPs with a high affinity to self MHC:peptide complexes are negatively selected and are marked as being autoreactive, undergoing apoptosis [180–182]. Negative selection is critical in checking autoimmunity [183, 184]. Further development of DP thymocytes to either

$CD4^+$ or $CD8^+$ SP requires the switching off of one of the co-receptors and is mediated in part by TCR signal duration and strength [185]. DP thymocytes are much more sensitive to the MHC:peptide complex than peripheral effector cells [179, 185]; however, the signaling cascades are very similar [186], and act in positive and negative feedback regulatory loops [179].

In the initial commitment from CLPs to T-cell fate, many epigenetic changes also take place along with Notch signaling or in response to it. For example, promoters of the B-cell lineage-determining factors Pax5 and Ebf1 are heavily methylated at H3K27(me3). Additionally, the "bivalent" *Cebpa* promoter discussed earlier remains bivalent, ensuring that CEBP expression specific for macrophages remains checked. Similarly, key erythroid genes such as Gata1 and EpoR are silenced with and without H3K27me3, respectively, in thymic lineages.

At early stages of T-cell development, in response to thymic Notch signaling, three T-cell fate regulators are upregulated, namely TCF1, GATA3, and Bcl11b [172] (Fig. 6). Bcl11b is sharply upregulated in the late DN2a stage and its deletion results in skewing of early T-cell progenitors to either myeloid or NK lineages [187–189]. Recent accessibility studies using DNase-seq and ATAC-seq in 8 different thymocyte stages revealed that the chromatin accessibility dramatically changes from DN2 to DN3 stages, and this was correlated with high expression of Bcl11b [190]. TCF1 (encoded by *Tcf7*) is known to be an important regulator in T-cell development [191, 192]. *Tcf7*, along with *Hes*1 and *Gata*3, is directly induced by Notch-DLL4 signaling, while *Tcf7* is targeted by the Notch-RBPJ transcription factor complex [192, 193]. TCF1 plays various roles in cell identity and lineage commitment in conjunction with other transcription factors such as HEB1 [194], LEF1 [195], and Runx3 [196]. A recent study using ATAC-seq showed that upon expression of TCF1 the chromatin landscape undergoes genome-wide alterations [197]. DNA methylation is also critical for proper T-cell development since deletion of DNMT1 results in almost complete loss of DP cells [198–200]. Furthermore, early development is accompanied by demethylation of many genes encoding CD3, Runx3, Lef1, Rorc, and Lck [198].

CD4/CD8 lineage commitment: The epigenetic contribution

Lineage choice of CD4 or CD8 cells is a long-standing conundrum, and is determined by a series of selection processes from the DP stage of T-cell development (Fig. 6). The prime regulator of CD4 differentiation is ThPOK, originally identified as a CD4 lineage commitment factor, since mice deficient in ThPOK no longer produce CD4 committed cells, but have normal CD8 cell numbers [201]. ThPOK is both necessary and sufficient for CD4 lineage commitment, as ectopic expression of ThPOK in thymocytes results in fate reversal of MHC class-I cells to MHC class-II expressing CD4 cells [201]. The *Thpok* gene itself is regulated via its proximal and distal enhancers [202]. The global chromatin organizer Special AT-rich Binding Protein 1 (SATB1) is highly enriched in thymic

T cells from the DP stage onward and directly regulates *Thpok*. Deletion of SATB1 in DP thymocytes results in impaired CD4 lineage commitment [203, 204]. Mechanistically, SATB1 binds to its enhancer and activates the gene expression, seemingly by changing the methylation status of histones. Another important role played by both ThPOK and SATB1 is in the positive regulation of the CD4 co-receptor [202]. ThPOK binds to an upstream "silencer" region of the *Cd4* locus and antagonizes its repressive nature, possibly by antagonizing the repressive complexes that bind nearby [205].

The Runt domain containing TFs Runx1/3 plays major roles in CD8 lineage commitment. During DP to CD8 transition in the thymus, Runx3 is actively produced [206]. Deletion of both Runx3 and Runx1 abolishes the development of CD8 cells [207]. Besides, Runx3 interacts with the TCF1-LEF1 complex which drives CD4 lineage commitment, resulting in suppression of various CD4 specific genes and induction of the CD8 specific transcription program [196].

Natural regulatory T cells (Tregs) develop in the thymus with the induction of FOXP3 lineage transcription factor [208, 209]. During CD4 lineage commitment, some CD4 cells receive stronger TCR signals, and begin to express CD25 and GITR (TNF family member) to become Treg precursors [210, 211] (Fig. 6). Treg precursors are associated with a change in the epigenetic landscape repressing effector CD4 specific genes, and upregulating genes important for regulatory function such as demethylation of cytotoxic T-lymphocyte associated protein 4 (CTLA-4), IL2R, and IL-2 [212]. A recent study characterized the enhancer landscape in Treg precursors and effector T helper cells by acetylation at H3K27ac [211]. Accordingly, 66 Treg specific super-enhancers were shown to have permissive H3K27ac, and this correlated with the global ATAC seq data in these cell types. Furthermore, SATB1 was shown to be an upstream regulator of these super-enhancers. SATB1-deficient Treg precursors lose the ability to give rise to mature Tregs [211].

Terminal differentiation and function of T cells in the periphery

In the periphery, CD4 T cells differentiate into various helper subsets categorized as TH1, TH2, TH17, T_{FH} cells, and induced regulatory T cells (iTregs) depending on the type of pathogen encountered and the cytokines present in the microenvironment. In the following section, we describe the epigenetic modifications associated with CD4 and CD8 T cells and key transcription factors that together modulate the proliferation, differentiation, and plasticity of CD4 and CD8 T cells.

Activation of naïve CD4 cells induces dramatic changes in their chromatin state. For example, activation-induced nuclear factor of activated T cells (NFAT)1 and AP1 (FOS/JUN and associated proteins) complex, along with ETS transcription factors, open up several sites in the chromatin [213], with an observed quantitative increase in acetylation (H3K27) at AP1 bound regions [214].

Differentiation of CD4 T cells into specific T helper subtypes is characterized by the expression of lineage-specific transcription factors, production of key cytokines, and is regulated by numerous epigenetic alterations [200–202]. Th1 cells secrete cytokines INFγ and TNF, Th2 cells secrete cytokines IL-4, IL-5, and IL-13 [215], and Th17 cells secrete IL-17A, IL-17F, and IL-22, among others. Lineage-specific transcription factors include T-bet and STAT1 for Th1 subtype, and GATA3 and STAT6, which shapes the Th2 lineage. GATA3 is induced at least in part by activation-specific TF SATB1 [216]. The Th17 subtype, on the other hand, requires retinoic acid-related orphan receptor RORγT and STAT3 [215]. Lineage-specific distal regulatory elements or enhancers also play a key role in T helper cell polarization [217]. Genome-wide binding analysis of H3K4me3 and H3K27me3 in helper subtypes Th1, Th2, and Th17 revealed that the genes encoding the aforementioned factors carry the permissive H3K4me3 mark only in their respective lineages, but are deposited with the repressive H3K27me3 mark in opposing lineages [218]. DNA methylation is another contributing modification in polarization of naïve CD4 cells to terminally differentiated cells [118, 219]. The promoter of the gene *Ifng*-encoding IFN-γ is heavily methylated in unpolarized CD4 T cells, but is actively demethylated only upon Th1 polarization [220]. Similarly, the genes encoding IL-4 and IL-13, specific to Th2 cells, and IL-17 and RORγt, specific to Th17 cells, are demethylated only in their corresponding lineage [221]. Further, the methylation of CNS2 intronic enhancer of Foxp3 in non-Treg cells ensures suppression of Treg gene expression profile in conventional CD4 T cells [118]. Conversely, this locus is actively demethylated by TET family proteins, as well as primed for expression by MLL-4 mediated H3K4me3 marks in Tregs [222, 223]. Moreover, loci of Treg-specific genes such as *Ctla4* and *Il2ra* are largely demethylated in regulatory T subsets.

Epigenetic changes play a central role in promoting the cytolytic property of activated CD8 T cells. The locus of *Gzmb* encoding granzyme B—a major cytolytic molecule—has heavy deposition of repressive H3K27me3 prior to activation, but this is actively removed upon activation and undergoes permissive H3K9ac and H3K4me3 modifications [224]. Additionally, IFN-γ produced by activated CD8 T cells is also regulated in a similar fashion [225].

Epigenetic changes in T-cell plasticity and memory

T-cell responses are highly dynamic and both the activation and polarization of T cells are under tight control of epigenetic changes for temporally correct and cell type-specific gene expression. HDAC1 and HDAC2 maintain commitment to CD4 lineage by repressing the gene program for CD8 effector cells [226]. Similarly, the histone methyltransferase SUV39H1 maintains Th2 stability by depositing repressive H3K9me on Th1-specific genes such as the IFN-γ encoding locus [227]. Furthermore, it was shown through early studies of HAT and HDAC ChIP-seq that there is a constant cycling of HDACs, on active genes in human CD4 T cells, suggesting an instant switch mechanism in T cells [218]. In addition, the *Tbx21* locus encoding the T-bet transcription factor is deposited with H3K4me3 only in Th1 subtypes, whereas there is bivalency at the same locus in other subsets with dual H3K4me3 and H3K27me3 marks, hinting at some degree of plasticity between the helper subsets [218].

The memory cells that are formed after primary antigen exposure also involve certain epigenetic modifications, allowing for a heightened and faster response upon reencounter with the same antigen [228]. Following activation-induced differentiation of naïve cells into various subtypes, these polarized helper cells secrete their characteristic cytokines due to permissive epigenetic marks such as H3K4me3 and H3K9ac. After clearance of infection and formation of memory cells, these active marks are not erased, and the chromatin remains in a permissive state, thereby ensuring a rapid response upon reexposure to the same antigen [229]. Another example of epigenetic memory is the presence of sustained activation marks and the presence of stably docked RNA pol-II on the IFN-γ promoter of CD8 cells until reactivation whereupon there is a surge of IFN-γ expression [230].

Gene regulation by long-distance interactions

Immune cells in the blood originate from a single hematopoietic stem cell precursor that undergoes a tightly regulated process of development and differentiation initially into the lymphoid or myeloid lineages and terminating into highly differentiated immune cell subtypes. This sequence of development involves lineage-specific transcriptional induction of molecules coupled in part with epigenetic regulations of which DNA methylation and histone modifications play an important role. In response to external stimulus, induction of transcription factors mediate chromatin remodeling and/or looping of the chromatin around a control region (Fig. 7A), resulting in configuration and reconfiguration of chromatin and ultimately stabilization in the terminally differentiated mature cell. This process involves cooperation of the promoter with more distal regulatory elements such as enhancers, insulators, silencers, and locus control regions that are frequently located ∼50 bases upstream or downstream of the gene.

Understanding the degree of conservation of the genome-wide epigenetic landscape of a cell (epigenome) [225] across species in the context of the primary DNA sequence is one way to understand epigenome evolution and to understand how changes in epigenetic marks including differential chromatin states contribute to differences in gene expression between species. Cain et al. estimated that changes in H3K4me3 levels between species could explain as much as 7% of the differences in gene expression profiles [224]. Other studies have indicated that the DNA methylation patterns for certain genomic elements that correspond to regulatory DNA sequences are conserved between vertebrates and plants [231–234], as are the location of nonmethylated CpG islands in vertebrate gene promoters [235]. Additionally, there is a greater difference in methylation

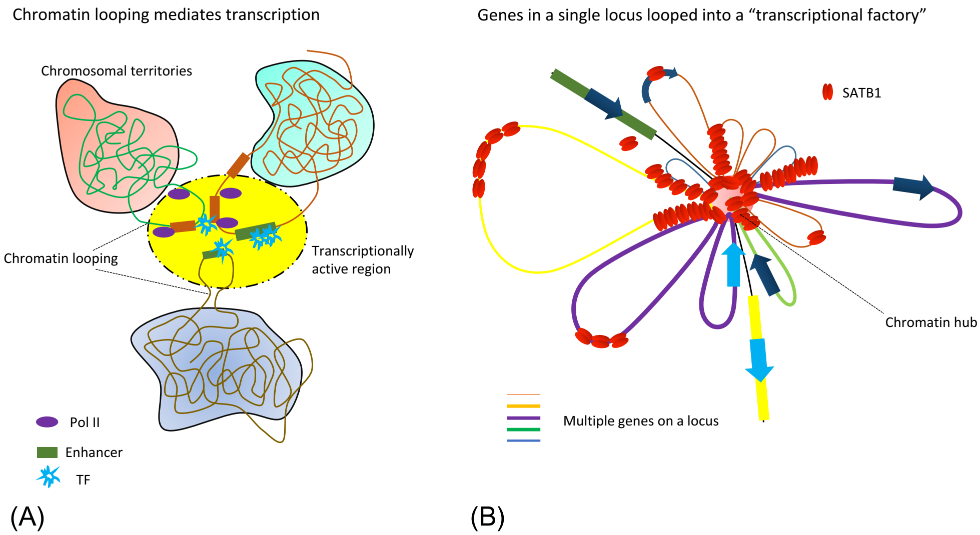

Chromatin looping mediates transcription
Chromosomal territories
Chromatin looping
Transcriptionally active region
Pol II
Enhancer
TF
(A)
Genes in a single locus looped into a "transcriptional factory"
SATB1
Chromatin hub
Multiple genes on a locus
(B)

Fig. 7 Transcriptional control by long-range DNA looping. (A) Chromosomes occupy specific "chromosomal territories" (CTs) in the nucleus. At the periphery, most of the chromatin exists in a repressed state in localized CTs, from which occurs distinct looping or chromatin extrusion events into the transcriptionally active region. The regulatory regions on the looped-out chromatin such as enhancers act as the docking sites for transcription factor and Pol II, which then mediate efficient transcription of genes which are brought closer due to long-range interactions. (B) Chromatin at a single locus—here the Th2 cytokine cluster is taken as an example. SATB1, a T-cell enriched TF and chromatin organizer, extensively loops multiple gene regulatory regions onto a single "chromatin hub" to mediate much tighter transcriptional control. SATB1 dynamically represses or activates different sets of genes at the MHC-I locus depending on the various looping events.

patterns between tissues of the same species than in the same tissue across species [236]. A recent study has compared the methylomes of three tissues across three species, namely rats, humans, and mice, and has shown up to 37% of rat tissue-specific differentially methylated regions (tsDMR) that are epigenetically conserved in humans and mice. These regions are associated with conservation of other active epigenetic modifications including histone modifications and are associated with conservation of the primary genomic sequence, which corresponds to active enhancers and/or promoters. They further showed that the genomic sequences were enriched for motifs corresponding to binding sites of transcription factors associated with the specific tissue, suggesting that the primary genomic sequence along with maintenance/turnover of transcription factors can drive conserved epigenetic states [237].

A number of studies on the regulation of IL-4 and IFN-γ loci in the immune system have revealed that *cis*-acting regulatory regions are frequently conserved across mammalian species. Shnyreva et al. identified two distal noncoding elements conserved between mice, rats, and humans, namely CNS1 and CNS2, that positively regulated IFN-γ production in T cells. These regions were associated with active histone acetylation that were T-bet dependent [220, 238]. Both T-bet and GATA-3 lineage transcription factors promote cytokine production in part by maintaining histone hyperacetylation, thereby facilitating accessibility to the transcriptional machinery [239, 240]. Long-range histone acetylation of the IFN-γ gene is an essential feature of T-cell differentiation [241]. The distal conserved element CNS-22 has also been shown to recruit transcription factors T-bet, Runx3, NF-kB, and STAT4 which regulate IFN-γ transcription in Th1 cells and that deletion of this element impaired IFN-γ expression through loss of an acute requirement of histone acetylation [242].

Loots et al. identified a distally located regulatory region, highly conserved in mice and humans named CNS-1 in the IL-4/IL-13 locus, that acted as an IL-4 enhancer. Deletion of this element impaired production of IL-4, IL-13, and IL-5 cytokines, suggesting its coordinated regulatory role at the Th2 cytokine locus [243]. Both Th1 and Th2 cells exhibit a looped chromatin conformation for lineage-specific cytokines. The Th2 cytokines IL-4, IL-13, and IL-5, although having their promoters situated a distance apart, are frequently expressed coordinately in part due to looping of promoters around a locus control region residing in the RAD50 gene body [244–248].

Chromatin looping in the Th2 cytokine locus is greatly facilitated by the SATB1 protein in Th2 cell type-specific manner (Fig. 7B). SATB1 functions as a global chromatin organizer and transcription factor [249, 250]. SATB1, based on its location (usually near enhancer sequences), interaction with different cofactors which are part of chromatin remodeling complexes, and ability to tether DNA elements, orchestrates the higher-order chromatin organization and tissue-specific gene transcription over long distances in the genome.

Upon Th2 activation, SATB1 promotes a higher-order transcriptionally active chromatin configuration wherein the chromatin within the 200 kb Th2 cytokine locus is folded into numerous small loops linked to SATB1 at its base. Of the nine SATB1 binding sites, four are located in the intronic region of the RAD50 gene body [250].

Using RNA interference, the same group further demonstrated that SATB1 was required for this folded looped chromatin structure, for c-maf induction, and for transcription of IL-4, IL-5, and IL-13 Th2 cytokines. Several Th2 factors were recruited to this region upon Th2 activation, including GATA-3, STAT-6, c-Maf, the chromatin-remodeling enzyme Brg1, and RNA polymerase II coupled with histone H3 acetylation [250]. SATB1 has been shown to interact with various cofactors and chromatin modifiers including β-catenin, p300, HDAC1, PCAF, and CtBP1, and binds to transcription factor loci such as GATA-3, that together regulate the epigenetic state and consequent gene transcription during T-cell development, T-cell activation, and T helper cell differentiation [206, 251–253].

Regulatory T cell (Treg)-specific genes were found to be associated with super-enhancers, which are by definition genomic regions with dense clustering of enhancers that were located near Treg signature genes such as Foxp3, Ctla4, IL2ra, and Ikzf2. These regions are associated with a higher degree of permissive epigenetic states, first observed at the thymic Treg precursor stage, and are also associated with transcription factors contributing to Treg cell function such as Foxp3, Runx1, Bcl11b, Ets1, and CREB [211]. SATB1 was shown to bind sequentially to different enhancers at different stages of Treg cell development, beginning from closed chromatin at the double positive stage to an active open chromatin configuration, suggesting its role in chromatin looping at these regulatory sequences. Development of thymic regulatory T cells was dependent on binding of SATB1 to regulatory T-cell super-enhancers, resulting in an open chromatin configuration through H3K27 acetylation promoting induction of Foxp3 and regulatory T-cell signature genes [211].

Similarly, SATB1 is also required for appropriate development of CD4 and CD8 T cells by binding to and regulating enhancers, activating genes of lineage-specifying factors including ThPOK, Runx3, CD4, and CD8 [202].

Concluding remarks and future perspective

Epigenetic modifications have a tremendous impact on the extent and diversity of transcriptional dynamics in cell development and function. Although many recent global accessibility and ChIP-seq studies have provided us clues for genome-wide epigenetic modulations in many cell types, there is remarkable scope for linking these changes to different transcription factors—which are only now being studied in the chromatin context. Presently, the information available on the epigenetic regulation of various key immune cells is primarily limited by the rarity of a population. Furthermore, the immune subtypes are classified using a relatively lesser number of surface markers, for example CD4 or CD8 cells. Recent single cell next-generation sequencing studies have revealed the existence of multiple distinct populations based on their transcriptomic profiles, and some new cell type-specificity markers may become significant. The recent advancements in high-throughput and single cell data generation and analysis such as ATAC-seq, CITE-seq, Hi-C, and methylated DNA immunoprecipitation-seq (MeDIP) are promising in providing a clearer picture of the dynamic epigenetic changes pertaining to each cell state.

Acknowledgments

This work was supported by a grant from the Unit of Excellence (BT/MED/30/SP11288/2015) Program of the Department of Biotechnology (DBT), Government of India, and Institutional support to SG. AM is supported by a fellowship from the CSIR, Government of India. AD is supported by the Center of Excellence in Epigenetics Program (BT/COE/34/SP17426/2016) of DBT awarded to SG.

References

[1] Allis CD, Jenuwein T. The molecular hallmarks of epigenetic control. Nat Rev Genet 2016;17(8):487–500.

[2] Luger K, Mader AW, Richmond RK, Sargent DF, Richmond TJ. Crystal structure of the nucleosome core particle at 2.8 Å resolution. Nature 1997;389:251–60.

[3] Handy DE, Castro R, Loscalzo J. Epigenetic modifications: basic mechanisms and role in cardiovascular disease. Circulation 2011;123(19):2145–56.

[4] Bird AP. CpG-rich islands and the function of DNA methylation. Nature 1986;321:209–13.

[5] Illingworth RS, Bird AP. CpG islands—'a rough guide'. FEBS Lett 2009;583:1713–20.

[6] Weber M, Hellmann I, Stadler MB, Ramos L, Paabo S, Rebhan M, Schubeler D. Distribution, silencing potential and evolutionary impact of promoter DNA methylation in the human genome. Nat Genet 2007;39:457–66.

[7] Nomura J, Hisatsune A, Miyata T, Isohama Y. The role of CpG methylation in cell type-specific expression of the aquaporin-5 gene. Biochem Biophys Res Commun 2007;353:1017–22.

[8] Li E, Bestor TH, Jaenisch R. Targeted mutation of the DNA methyltransferase gene results in embryonic lethality. Cell 1992;69(6):915–26.

[9] Okano M, Bell DW, Haber DA, Li E. DNA methyltransferases Dnmt3a and Dnmt3b are essential for de novo methylation and mammalian development. Cell 1999;99(3):247–57.

[10] Brownell JE, Zhou J, Ranalli T, Kobayashi R, Edmondson DG, Roth SY, et al. Tetrahymena histone acetyltransferase A: a homolog to yeast Gcn5p linking histone acetylation to gene activation. Cell 1996;84(6):843–51.

[11] Kuo MH, Brownell JE, Sobel RE, Ranalli TA, Cook RG, Edmondson DG, et al. Transcription-linked acetylation by Gcn5p of histones H3 and H4 at specific lysines. Nature 1996;383(6597):269–72.

[12] Fischle W, Wang Y, Allis CD. Binary switches and modification cassettes in histone biology and beyond. Nature 2003;425:475–9.

[13] Lawrence M, Daujat S, Schneider R. Lateral thinking: how histone modifications regulate gene expression. Trends Genet 2016;32(1):42–56.

[14] Strahl BD, Allis CD. The language of covalent histone modifications. Nature 2000;403:41–5.

[15] Margueron R, Reinberg D. The Polycomb complex PRC2 and its mark in life. Nature 2011;469:343–9.

[16] Gaydos LJ, Wang W, Strome S. Gene repression. H3K27me and PRC2 transmit a memory of repression across generations and during development. Science 2014;345:1515–8.

[17] Czermin B, et al. Drosophila enhancer of zeste/ESC complexes have a histone H3 methyltransferase activity that marks chromosomal Polycomb sites. Cell 2002;111:185–96.

[18] Melcher M, et al. Structure-function analysis of SUV39H1 reveals a dominant role in heterochromatin organization, chromosome segregation, and mitotic progression. Mol Cell Biol 2000;20:3728–41.

[19] Rea S, et al. Regulation of chromatin structure by site-specific histone H3 methyltransferases. Nature 2000;406:593–9.

[20] Tachibana M, Sugimoto K, Fukushima T, Shinkai Y. SET domain-containing protein, G9a, is a novel lysine-preferring mammalian histone methyltransferase with hyperactivity and specific selectivity to lysines 9 and 27 of histone H3. J Biol Chem 2001;276:25309–17.

[21] Sanders SL, Portoso M, Mata J, Bähler J, Allshire RC, Kouzarides T. Methylation of histone H4 lysine 20 controls recruitment of Crb2 to sites of DNA damage. Cell 2004;119(24):603–14.

[22] Nguyen AT, Zhang Y. The diverse functions of Dot1 and H3K79 methylation. Genes Dev 2011;25(13):1345–58.

[23] Orstavik KH. X chromosome inactivation in clinical practice. Hum Genet 2009;126:363–73.

[24] Wagschal A, Sutherland HG, Woodfine K, Henckel A, Chebli K, Schulz R, Oakey RJ, Bickmore WA, Feil R. G9a histone methyltransferase contributes to imprinting in the mouse placenta. Mol Cell Biol 2008;28:1104–13.

[25] Cedar H, Bergman Y. Linking DNA methylation and histone modification: patterns and paradigms. Nat Rev Genet 2009;10: pages 295–304.
[26] Feldman N, et al. G9a-mediated irreversible epigenetic inactivation of Oct-3/4 during early embryogenesis. Nat Cell Biol 2006;8:188–94.
[27] Blackledge NP, Zhou JC, Tolstorukov MY, Farcas AM, Park PJ, Klose RJ. CpG islands recruit a histone H3 lysine 36 demethylase. Mol Cell 2010;38(2):179–90.
[28] Thomson JP, Skene PJ, Selfridge J, Clouaire T, Guy J, Webb S, et al. CpG islands influence chromatin structure via the CpG-binding protein Cfp1. Nature 2010;464(7291):1082–6.
[29] Ooi SK, et al. DNMT3L connects unmethylated lysine 4 of histone H3 to de novo methylation of DNA. Nature 2007;448:714–7.
[30] Ho L, Crabtree GR. Chromatin remodelling during development. Nature 2010;463:474–84.
[31] Clapier CR, Cairns BR. In: Workman JL, Abmayr SM, editors. Fundamentals of chromatin. Springer; 2014. p. 69–146.
[32] Clapier CR, Iwasa J, Cairns BR, Craig L. Peterson mechanisms of action and regulation of ATP-dependent chromatin-remodel*ling* complexes. Nat Rev Mol Cell Biol 2017;18:407–22.
[33] Costa FF. Non-coding RNAs, epigenetics and complexity. Gene 2008;410:9–17.
[34] Amaral PP, Dinger ME, Mercer TR, Mattick JS. The eukaryotic genome as an RNA machine. Science 2008;319:1787–9.
[35] Ghildiyal M, Zamore PD. Small silencing RNAs: an expanding universe. Nat Rev Genet 2009;10: 94–108.
[36] Zhou X, Ren Y, Zhang J, Zhang C, Zhang K, Han L, Kong L, Wei J, Chen L, Yang J, et al. HOTAIR is a therapeutic target in glioblastoma. Oncotarget 2015;6:8353–65.
[37] Zhang K, Sun X, Zhou X, Han L, Chen L, Shi Z, Zhang A, Ye M, Wang Q, Liu C, et al. Long noncoding RNA HOTAIR promotes glioblastoma cell cycle progression in an EZH2 dependent manner. Oncotarget 2015;6:537–46.
[38] Rinn JL, Kertesz M, Wang JK, Squazzo SL, Xu X, Brugmann SA, Goodnough LH, Helms JA, Farnham PJ, Segal E, et al. Functional demarcation of active and silent chromatin domains in human HOX loci by noncoding RNAs. Cell 2007;129:1311–23.
[39] Huang XA, Yin H, Sweeney S, Raha D, Snyder M, Lin H. A major epigenetic programming mechanism guided by piRNAs. Dev Cell 2013;24:502–16.
[40] Buenrostro JD, Giresi PG, Zaba LC, Chang HY, Greenleaf WJ. Transposition of native chromatin for fast and sensitive epigenomic profiling of open chromatin, DNA-binding proteins and nucleosome position. Nat Methods 2013;10:1213–8.
[41] Neph S, Vierstra J, Stergachis AB, Reynolds AP, Haugen E, Vernot B, Thurman RE, John S, Sandstrom R, Johnson AK, Maurano MT, Humbert R, Rynes E, Wang H, Vong S, Lee K, Bates D, Diegel M, Roach V, Dunn D, Neri J, Schafer A, Hansen RS, Kutyavin T, Giste E, Weaver M, Canfield T, Sabo P, Zhang M, Balasundaram G, Byron R, MacCoss MJ, Akey JM, Bender MA, Groudine M, Kaul R, Stamatoyannopoulos JA. An expansive human regulatory lexicon encoded in transcription factor footprints. Nature 2012;489:83–90.
[42] Heinz S, Romanoski CE, Benner C, Glass CK. The selection and function of cell type-specific enhancers. Nat Rev Mol Cell Biol 2015;16:144–54.
[43] Bernstein BE, Mikkelsen TS, Xie X, Kamal M, Huebert DJ, Cuff J, et al. A bivalent chromatin structure marks key developmental genes in embryonic stem cells. Cell 2006;125(2):315–26.
[44] Sachs M, Onodera C, Blaschke K, Ebata KT, Song JS, Ramalho-Santos M. Bivalent chromatin marks developmental regulatory genes in the mouse embryonic germline in vivo. Cell Rep 2013;3(6): 1777–84.
[45] Roh TY, Cuddapah S, Cui K, Zhao K. The genomic landscape of histone modifications in human T cells. Proc Natl Acad Sci U S A 2006;103:15782–7.
[46] Pan G, Tian S, Nie J, Yang C, Ruotti V, Wei H, Jonsdottir GA, Stewart R, Thomson JA. Whole-genome analysis of histone h3 lysine 4 and lysine 27 methylation in human embryonic stem cells. Cell Stem Cell 2007;1:299–312.
[47] Martino D, Kesper DA, Amarasekera M, Harb H, Renz H, Prescott S. Epigenetics in immune development and in allergic and autoimmune diseases. J Reprod Immunol 2014;104–105:43–8.

[48] Broeske A-M, Vockentanz L, Kharazi S, Huska MR, Mancini E, Scheller M, Kuhl C, Enns A, Prinz M, Jaenisch R, Nerlov C, Leutz A, Andrade-Navarro MA, Jacobsen SEW, Rosenbauer F. DNA methylation protects hematopoietic stem cell multipotency from myeloerythroid restriction. Nat Genet 2009;41:1207–U69.

[49] Moran-Crusio K, Reavie L, Shih A, Abdel-Wahab O, Ndiaye-Lobry D, Lobry C, Figueroa ME, Vasanthakumar A, Patel J, Zhao X, Perna F, Pandey S, Madzo J, Song C, Dai Q, He C, Ibrahim S, Beran M, Zavadil J, Nimer SD, Melnick A, Godley LA, Aifantis I, Levine RL. Tet2 loss leads to increased hematopoietic stem cell selfrenewal and myeloid transformation. Cancer Cell 2011;20:11–24.

[50] Di Santo JP. Natural killer cell developmental pathways: a question of balance. Annu Rev Immunol 2006;24:257–86.

[51] Shi FD, Ljunggren HG, La Cava A, Van Kaer L. Organ-specific features of natural killer cells. Nat Rev Immunol 2011;11(10):658–71.

[52] Geiger TL, Joseph C. Sun development and maturation of natural killer cells. Curr Opin Immunol 2016;39:82–9.

[53] Orr MT, Lanier LL. Natural killer cell education and tolerance. Cell 2010;142:847–56.

[54] Lanier LL. Up on the tightrope: natural killer cell activation and inhibition. Nat Immunol 2008; 9(5):495–502.

[55] Schenk A, Bloch W, Zimmer P. Natural killer cells: an epigenetic perspective of development and regulation. Int J Mol Sci 2016;17(3):326.

[56] Cichocki F, Miller JS, Anderson SK, Bryceson YT. Epigenetic regulation of NK cell differentiation and effector functions. Front Immunol 2013;4:55.

[57] Santourlidis S, Graffmann N, Christ J, Uhrberg M. Lineage-specific transition of histone signatures in the killer cell Ig-like receptor locus from hematopoietic progenitor to NK cells. J Immunol 2008; 180:418–25.

[58] Gao X-N, Lin J, Wang L-L, Yu L. Demethylating treatment suppresses natural killer cell cytolytic activity. Mol Immunol 2009;46:2064–70.

[59] Ogbomo H, Michaelis M, Kreuter J, Doerr HW, Cinatl J. Histone deacetylase inhibitors suppress natural killer cell cytolytic activity. FEBS Lett 2007;581:1317–22.

[60] Santourlidis S, Trompeter H-I, Weinhold S, Eisermann B, Meyer KL, Wernet P, Uhrberg M. Crucial role of DNA methylation in determination of clonally distributed killer cell Ig-like receptor expression patterns in NK cells. J Immunol 2002;169:4253–61.

[61] Fernández-Sánchez A, Baragaño RA, Carvajal PR, Sanz AB, Ortiz A, Ortega F, Suárez-Álvarez B, López-Larrea C. DNA demethylation and histone H3K9 acetylation determine the active transcription of the NKG2D gene in human CD8+ T and NK cells. Epigenetics 2013;8:66–78.

[62] DeHart SL, Heikens MJ, Tsai S. Jagged2 promotes the development of natural killer cells and the establishment of functional natural killer cell lines. Blood 2005;105:3521–7.

[63] Jaleco AC, Neves H, Hooijberg E, Gameiro P, Clode N, Haury M, Henrique D, Parreira L. Differential effects of notch ligands Delta-1 and Jagged-1 in human lymphoid differentiation. J Exp Med 2001;194:991–1002.

[64] Gascoyne DM, Long E, Veiga-Fernandes H, de Boer J, Williams O, Seddon B, Coles M, Kioussis D, Brady HJM. The basic leucine zipper transcription factor E4BP4 is essential for natural killer cell development. Nat Immunol 2009;10:1118–24.

[65] Kamizono S, Duncan GS, Seidel MG, Morimoto A, Hamada K, Grosveld G, Akashi K, Lind EF, Haight JP, Ohashi PS, et al. Nfil3/E4bp4 is required for the development and maturation of NK cells in vivo. J Exp Med 2009;206:2977–86.

[66] Yokota Y, Mansouri A, Mori S, Sugawara S, Adachi S, Nishikawa S, Gruss P. Development of peripheral lymphoid organs and natural killer cells depends on the helix-loop-helix inhibitor Id2. Nature 1999;397:702–6.

[67] Kashiwada M, Pham N-LL, Pewe LL, Harty JT, Rothman PB. NFIL3/E4BP4 is a key transcription factor for CD8α+ dendritic cell development. Blood 2011;117:6193–7.

[68] Geiger TL, Abt MC, Gasteiger G, Firth MA, O'Connor MH, Geary CD, O'Sullivan TE, van den Brink MR, Pamer EG, Hanash AM, et al. Nfil3 is crucial for development of innate lymphoid cells and host protection against intestinal pathogens. J Exp Med 2014;211: 1723–31.

[69] Seillet C, Rankin LC, Groom JR, Mielke LA, Tellier J, Chopin M, Huntington ND, Belz GT, Carotta S. Nfil3 is required for the development of all innate lymphoid cell subsets. J Exp Med 2014; 211:1733–40.
[70] Kashiwada M, Levy DM, McKeag L, Murray K, Schröder AJ, Canfield SM, Traver G, Rothman PB. IL-4-induced transcription factor NFIL3/E4BP4 controls IgE class switching. Proc Natl Acad Sci U S A 2010;107:821–6.
[71] Yang M, Li D, Chang Z, Yang Z, Tian Z, Dong Z. PDK1 orchestrates early NK cell development through induction of E4BP4 expression and maintenance of IL-15 responsiveness. J Exp Med 2015;212:253–65.
[72] Schotte R, Dontje W, Nagasawa M, Yasuda Y, Bakker AQ, Spits H, Blom B. Synergy between IL-15 and Id2 promotes the expansion of human NK progenitor cells, which can be counteracted by the E protein HEB required to drive T cell development. J Immunol 2010;184:6670–9.
[73] Imada K, Bloom ET, Nakajima H, Horvath-Arcidiacono JA, Udy GB, Davey HW, Leonard WJ. Stat5b is essential for natural killer cell-mediated proliferation and cytolytic activity. J Exp Med 1998;188:2067–74.
[74] Eckelhart E, Warsch W, Zebedin E, Simma O, Stoiber D, Kolbe T, Rülicke T, Mueller M, Casanova E, Sexl V. A novel Ncr1-Cre mouse reveals the essential role of STAT5 for NK-cell survival and development. Blood 2011;117:1565–73.
[75] Guo Y, Maillard I, Chakraborti S, Rothenberg EV, Speck NA. Core binding factors are necessary for natural killer cell development and cooperate with notch signaling during T-cell specification. Blood 2008;112:480–92.
[76] Zhang J, Scordi I, Smyth MJ, Lichtenheld MG. Interleukin 2 receptor signaling regulates the perforin gene through signal transducer and activator of transcription (Stat)5 activation of two enhancers. J Exp Med 1999;190(9):1297–308.
[77] Hatton RD, Harrington LE, Luther RJ, Wakefield T, Janowski KM, Oliver JR, et al. A distal conserved sequence element controls Ifng gene expression by T cells and NK cells. Immunity 2006;25(5):717–29.
[78] Moro K, Yamada T, Tanabe M, Takeuchi T, Ikawa T, Kawamoto H, et al. Innate production of T(H) 2 cytokines by adipose tissue-associated c-Kit(+)Sca-1(+) lymphoid cells. Nature 2010;463(7280): 540–4.
[79] Neill DR, Wong SH, Bellosi A, Flynn RJ, Daly M, Langford TK, et al. Nuocytes represent a new innate effector leukocyte that mediates type-2 immunity. Nature 2010;464(7293):1367–70.
[80] Price AE, Liang HE, Sullivan BM, Reinhardt RL, Eisley CJ, Erle DJ, et al. Systemically dispersed innate IL-13-expressing cells in type 2 immunity. Proc Natl Acad Sci U S A 2010; 107(25):11489–94.
[81] Bernink JH, Peters CP, Munneke M, te Velde AA, Meijer SL, Weijer K, et al. Human type 1 innate lymphoid cells accumulate in inflamed mucosal tissues. Nat Immunol 2013;14(3):221–9.
[82] Halim TY, Krauss RH, Sun AC, Takei F. Lung natural helper cells are a critical source of Th2 cell-type cytokines in protease allergen-induced airway inflammation. Immunity 2012; 36(3):451–63.
[83] Klein Wolterink RG, Serafini N, van Nimwegen M, Vosshenrich CA, de Bruijn MJ, Fonseca Pereira D, et al. Essential, dose-dependent role for the transcription factor Gata3 in the development of IL-5+ and IL-13+ type 2 innate lymphoid cells. Proc Natl Acad Sci U S A 2013;110(25): 10240–5.
[84] Hoyler T, Klose CS, Souabni A, Turqueti-Neves A, Pfeifer D, Rawlins EL, et al. The transcription factor GATA-3 controls cell fate and maintenance of type 2 innate lymphoid cells. Immunity 2012;37 (4):634–48.
[85] Yagi R, Zhong C, Northrup DL, Yu F, Bouladoux N, Spencer S, et al. The transcription factor GATA3 is critical for the development of all IL-7Ralpha-expressing innate lymphoid cells. Immunity 2014;40(3):378–88.
[86] Yang Q, Monticelli LA, Saenz SA, Chi AW, Sonnenberg GF, Tang J, et al. T cell factor 1 is required for group 2 innate lymphoid cell generation. Immunity 2013;38(4):694–704.
[87] Mielke LA, Groom JR, Rankin LC, Seillet C, Masson F, Putoczki T, et al. TCF-1 controls ILC2 and NKp46+RORgammat+ innate lymphocyte differentiation and protection in intestinal inflammation. J Immunol 2013;191(8):4383–91.

[88] Spooner CJ, Lesch J, Yan D, Khan AA, Abbas A, Ramirez-Carrozzi V, et al. Specification of type 2 innate lymphocytes by the transcriptional determinant Gfi1. Nat Immunol 2013;14(12):1229–36.

[89] Amit I, Winter DR, Jung S. The role of the local environment and epigenetics in shaping macrophage identity and their effect on tissue homeostasis. Nat Immunol 2016;17:18–25.

[90] Gordon S, Martinez FO. Alternative activation of macrophages: mechanism and functions. Immunity 2010;32(5):593–604.

[91] Mosser DM, Edwards JP. Exploring the full spectrum of macrophage activation. Nat Rev Immunol 2008;8(12):958–69.

[92] Alvarez-Errico D, Vento-Tormo R, Sieweke M, Ballestar E. Epigenetic control of myeloid cell differentiation, identity and function. Nat Rev Immunol 2015;15(1):7–17.

[93] Lara-Astiaso D, et al. Immunogenetics. Chromatin state dynamics during blood formation. Science 2014;345:943–9.

[94] Sun D, Luo M, Jeong M, Rodriguez B, Xia Z, Hannah R, et al. Epigenomic profiling of young and aged HSCs reveals concerted changes during aging that reinforce self-renewal. Cell Stem Cell 2014;14(5):673–88.

[95] Heinz S, et al. Simple combinations of lineage-determining transcription factors prime cis-regulatory elements required for macrophage and B cell identities. Mol Cell 2010;38:576–89.

[96] Ghisletti S, et al. Identification and characterization of enhancers controlling the inflammatory gene expression program in macrophages. Immunity 2010;32:317–28.

[97] Okabe Y, Medzhitov R. Tissue-specific signals control reversible program of localization and functional polarization of macrophages. Cell 2014;157:832–44.

[98] Ostuni R, Natoli G. Lineages, cell types and functional states: a genomic view. Curr Opin Cell Biol 2013;25:759–64.

[99] de Groot AE, Pienta KJ. Epigenetic control of macrophage polarization: implications for targeting tumor-associated macrophages. Oncotarget 2018;9(29):20908–27.

[100] Ivashkiv LB. Epigenetic regulation of macrophage polarization and function. Trends Immunol 2013;34(5):216–23.

[101] Lawrence T, Natoli G. Transcriptional regulation of macrophage polarization: enabling diversity with identity. Nat Rev Immunol 2011;11:750–61.

[102] Biswas SK, Mantovani A. Macrophage plasticity and interaction with lymphocyte subsets: cancer as a paradigm. Nat Immunol 2010;11(10):889–96.

[103] Barish GD, Yu RT, Karunasiri M, Ocampo CB, Dixon J, Benner C, Dent AL, Tangirala RK, Evans RM. Bcl-6 and NF-kappaB cistromes mediate opposing regulation of the innate immune response. Genes Dev 2010;24:2760–5.

[104] De Santa F, Narang V, Yap ZH, Tusi BK, Burgold T, Austenaa L, Bucci G, Caganova M, Notarbartolo S, Casola S, Testa G, Sung WK, Wei CL, Natoli G. Jmjd3 contributes to the control of gene expression in LPS-activated macrophages. EMBO J 2009;28:3341–52.

[105] Escoubet-Lozach L, Benner C, Kaikkonen MU, Lozach J, Heinz S, Spann NJ, Crotti A, Stender J, Ghisletti S, Reichart D, Cheng CS, Luna R, Ludka C, Sasik R, Garcia-Bassets I, Hoffmann A, Subramaniam S, Hardiman G, Rosenfeld MG, Glass CK. Mechanisms establishing TLR4- responsive activation states of inflammatory response genes. PLoS Genet 2011;7:e1002401.

[106] Hargreaves DC, Horng T, Medzhitov R. Control of inducible gene expression by signal-dependent transcriptional elongation. Cell 2009;138:129–45.

[107] Ramirez-Carrozzi VR, Nazarian AA, Li CC, Gore SL, Sridharan R, Imbalzano AN, Smale ST. Selective and antagonistic functions of SWI/SNF and Mi-2beta nucleosome remodeling complexes during an inflammatory response. Genes Dev 2006;20:282–96.

[108] Jin F, Li Y, Ren B, Natarajan R. PU.1 and C/EBP(alpha) synergistically program distinct response to NF-kappaB activation through establishing monocyte specific enhancers. Proc Natl Acad Sci U S A 2011;108:5290–5.

[109] Pham TH, Benner C, Lichtinger M, Schwarzfischer L, Hu Y, Andreesen R, Chen W, Rehli M. Dynamic epigenetic enhancer signatures reveal key transcription factors associated with monocytic differentiation states. Blood 2012;119:e161–71.

[110] Glass CK, Saijo K. Nuclear receptor transrepression pathways that regulate inflammation in macrophages and T cells. Nat Rev Immunol 2010;10:365–76.

[111] Kruidenier L, Chung CW, Cheng Z, Liddle J, Che K, Joberty G, Bantscheff M, Bountra C, Bridges A, Diallo H, Eberhard D, Hutchinson S, Jones E, Katso R, Leveridge M, Mander PK, Mosley J, Ramirez-Molina C, Rowland P, Schofield CJ, Sheppard RJ, Smith JE, Swales C, Tanner R, Thomas P, Tumber A, Drewes G, Oppermann U, Patel DJ, Lee K, Wilson DM. A selective jumonji H3K27 demethylase inhibitor modulates the proinflammatory macrophage response. Nature 2012; 488:404–8.

[112] Zhu Y, van Essen D, Saccani S. Cell-type-specific control of enhancer activity by H3K9 trimethylation. Mol Cell 2012;46:408–23.

[113] Fang TC, Schaefer U, Mecklenbrauker I, Stienen A, Dewell S, Chen MS, Rioja I, Parravicini V, Prinjha RK, Chandwani R, MR MD, Lee K, Rice CM, Tarakhovsky A. Histone H3 lysine 9 di-methylation as an epigenetic signature of the interferon response. J Exp Med 2012;209:661–9.

[114] Stender JD, Pascual G, Liu W, Kaikkonen MU, Do K, Spann NJ, Boutros M, Perrimon N, Rosenfeld MG, Glass CK. Control of proinflammatory gene programs by regulated trimethylation and demethylation of histone H4K20. Mol Cell 2012;48:28–38.

[115] Ramirez-Carrozzi VR, Braas D, Bhatt DM, Cheng CS, Hong C, Doty KR, Black JC, Hoffmann A, Carey M, Smale ST. A unifying model for the selective regulation of inducible transcription by CpG islands and nucleosome remodeling. Cell 2009;138:114–28.

[116] Satoh T, Takeuchi O, Vandenbon A, Yasuda K, Tanaka Y, Kumagai Y, Miyake T, Matsushita K, Okazaki T, Saitoh T, Honma K, Matsuyama T, Yui K, Tsujimura T, Standley DM, Nakanishi K, Nakai K, Akira S. The Jmjd3-Irf4 axis regulates M2 macrophage polarization and host responses against helminth infection. Nat Immunol 2010;11:936–44.

[117] Mullican SE, Gaddis CA, Alenghat T, Nair MG, Giacomin PR, Everett LJ, Feng D, Steger DJ, Schug J, Artis D, Lazar MA. Histone deacetylase 3 is an epigenomic brake in macrophage alternative activation. Genes Dev 2011;25:2480–8.

[118] Wilson CB, Rowell E, Sekimata M. Epigenetic control of T-helper-cell differentiation. Nat Rev Immunol 2009;9(2):91–105.

[119] Miller JC, Brown BD, Shay T, Gautier EL, Jojic V, Cohain A, Pandey G, Leboeuf M, Elpek KG, Helft J, Hashimoto D, Chow A, Price J, Greter M, Bogunovic M, Bellemare-Pelletier A, Frenette PS, Randolph GJ, Turley SJ, Merad M, Immunological Genome Consortium. Deciphering the transcriptional network of the dendritic cell lineage. Nat Immunol 2012;13:888–99.

[120] Watowich SS, Liu YJ. Mechanisms regulating dendritic cell specification and development. Immunol Rev 2010;238:76–92.

[121] Collin M, McGovern N, Haniffa M. Human dendritic cell subsets. Immunology 2013;140:22–30.

[122] Colonna M, Trinchieri G, Liu YJ. Plasmacytoid dendritic cells in immunity. Nat Immunol 2004; 5:1219–26.

[123] Shortman K, Naik SH. Steady-state and inflammatory dendritic-cell development. Nat Rev Immunol 2007;7:19–30.

[124] Steinman RM, Banchereau J. Taking dendritic cells into medicine. Nature 2007;449:419–26.

[125] Graf T, Enver T. Forcing cells to change lineages. Nature 2009;462:587–94.

[126] Lin Q, Chauvistré H, Costa IG, Gusmao EG, Mitzka S, Hänzelmann S, Baying B, Klisch T, Moriggl R, Hennuy B, Smeets H, Hoffmann K, Benes V, Seré K, Zenke M. Epigenetic program and transcription factor circuitry of dendritic cell development. Nucleic Acids Res 2015;43: 9680–93.

[127] Yu H, Pardoll D, Jove R. STATs in cancer inflammation and immunity: a leading role for STAT3. Nat Rev Cancer 2009;9:798–809.

[128] Laouar Y, Welte T, Fu XY, Flavell RA. STAT3 is required for Flt3L-dependent dendritic cell differentiation. Immunity 2003;19:903–12.

[129] Karsunky H, Merad M, Cozzio A, Weissman IL, Manz MG. Flt3 ligand regulates dendritic cell development from Flt3+ lymphoid and myeloid-committed progenitors to Flt3+ dendritic cells in vivo. J Exp Med 2003;198:305–13.

[130] Anderson KL, Perkin H, Surh CD, Venturini S, Maki RA, Torbett BE. Transcription factor PU.1 is necessary for development of thymic and myeloid progenitor-derived dendritic cells. J Immunol 2000;164:1855–61.

[131] Tian Y, Meng L, Zhang Y. Epigenetic regulation of dendritic cell development and function. Cancer J 2017;23(5):302–7.

[132] Paul F, Arkin Y, Giladi A, Jaitin DA, Kenigsberg E, Keren-Shaul H, Winter D, Lara-Astiaso D, Gury M, Weiner A, David E, Cohen N, Lauridsen FK, Haas S, Schlitzer A, Mildner A, Ginhoux F, Jung S, Trumpp A, Porse BT, Tanay A, Amit I. Transcriptional heterogeneity and lineage commitment in myeloid progenitors. Cell 2015;163:1663–77.

[133] Clark MR, Mandal M, Ochiai K, Singh H. Orchestrating B cell lymphopoiesis through interplay of IL-7 receptor and pre-B cell receptor signalling. Nat Rev Immunol 2014;14:69–80.

[134] Bao Y, Cao X. The immune potential and immunopathology of cytokine-producing B cell subsets: a comprehensive review. J Autoimmun 2014;55:10–23.

[135] Cooper MD. The early history of B cells. Nat Rev Immunol 2015;15:191–7.

[136] Reth M, Nielsen P. Signaling circuits in early B-cell development. Adv Immunol 2014;122:129–75.

[137] Shaffer AL, Yu X, He Y, Boldrick J, Chan EP, Staudt LM. BCL-6 represses genes that function in lymphocyte differentiation, inflammation and cell cycle control. Immunity 2000;13:199–212.

[138] Treiber N, Treiber T, Zocher G, Grosschedl R. Structure of an Ebf1:DNA complex reveals unusual DNA recognition and structural homology with Rel proteins. Genes Dev 2010;24(20):2270–5.

[139] Lin YC, Jhunjhunwala S, Benner C, Heinz S, Welinder E, Mansson R, Sigvardsson M, Hagman J, Espinoza CA, Dutkowski J, Ideker T, Glass CK, Murre C. A global network of transcription factors, involving E2A, EBF1 and Foxo1, that orchestrates B cell fate. Nat Immunol 2010;11(7):635–43.

[140] Nutt SL, Heavey B, Rolink AG, Busslinger M. Commitment to the B-lymphoid lineage depends on the transcription factor Pax5. Nature 1999;401(6753):556–62.

[141] Maier H, Ostraat R, Gao H, Fields S, Shinton SA, Medina KL, Ikawa T, Murre C, Singh H, Hardy RR, Hagman J. Early B cell factor cooperates with Runx1 and mediates epigenetic changes associated with mb-1 transcription. Nat Immunol 2004;5(10):1069–77.

[142] Gao H, Lukin K, Ramirez J, Fields S, Lopez D, Hagman J. Opposing effects of SWI/SNF and Mi-2/NuRD chromatin remodeling complexes on epigenetic reprogramming by EBF and Pax5. Proc Natl Acad Sci U S A 2009;106(27):11258–63.

[143] Medvedovic J, Ebert A, Tagoh H, Busslinger M. Pax5: a master regulator of B cell development and leukemogenesis. Adv Immunol 2011;111:179–206.

[144] Zouali M. The epigenetic landscape of B lymphocyte tolerance to self. FEBS Lett 2013;587 (13):2067–73.

[145] Cho YW, Hong T, Hong S, Guo H, Yu H, Kim D, et al. PTIP associates with MLL3- and MLL4-containing histone H3 lysine 4 methyltransferase complex. J Biol Chem 2007;282(28):20395–406.

[146] McManus S, Ebert A, Salvagiotto G, Medvedovic J, Sun Q, Tamir I, Jaritz M, Tagoh H, Busslinger M. The transcription factor Pax5 regulates its target genes by recruiting chromatin-modifying proteins in committed B cells. EMBO J 2011;30:2388–404.

[147] Bao Y, Cao X. Epigenetic control of B cell development and B-cell-related immune disorders. Clin Rev Allergy Immunol 2016;50(3):301–11.

[148] Hatta M, Cirillo LA. Chromatin opening and stable perturbation of core histone:DNA contacts by FoxO1. J Biol Chem 2007;282:35583–93.

[149] Zhou B, Wang S, Mayr C, Bartel DP, Lodish HF. miR-150, a microRNA expressed in mature B and T cells, blocks early B cell development when expressed prematurely. Proc Natl Acad Sci U S A 2007;104:7080–5.

[150] Xiao C, Calado DP, Galler G, et al. MiR-150 controls B cell differentiation by targeting the transcription factor c-Myb. Cell 2007;131:146–59.

[151] Koralov SB, Muljo SA, Galler GR, et al. Dicer ablation affects antibody diversity and cell survival in the B lymphocyte lineage. Cell 2008;132:860–74.

[152] Xu CR, Feeney AJ. The epigenetic profile of Ig genes is dynamically regulated during B cell differentiation and is modulated by pre-B cell receptor signaling. J Immunol 2009;182(3):1362–9.

[153] McMurry MT, Krangel MS. A role for histone acetylation in the developmental regulation of VDJ recombination. Science 2000;287(5452):495–8.
[154] Johnson K, Angelin-Duclos C, Park S, Calame KL. Changes in histone acetylation are associated with differences in accessibility of V(H) gene segments to V-DJ recombination during B-cell ontogeny and development. Mol Cell Biol 2003;23(7):2438–50.
[155] Morshead KB, Ciccone DN, Taverna SD, Allis CD, Oettinger MA. Antigen receptor loci poised for V(D)J rearrangement are broadly associated with BRG1 and flanked by peaks of histone H3 dimethylated at lysine 4. Proc Natl Acad Sci U S A 2003;100(20):11577–82.
[156] Perkins EJ, Kee BL, Ramsden DA. Histone 3 lysine 4 methylation during the pre-B to immature B-cell transition. Nucleic Acids Res 2004;32(6):1942–7.
[157] Goldmit M, Ji Y, Skok J, Roldan E, Jung S, Cedar H, et al. Epigenetic ontogeny of the Igk locus during B cell development. Nat Immunol 2005;6(2):198–203.
[158] Shimazaki N, Tsai AG, Lieber MR. H3K4me3 stimulates the V(D)J RAG complex for both nicking and hairpinning in trans in addition to tethering in cis: implications for translocations. Mol Cell 2009;34(5):535–44.
[159] Nakase H, Takahama Y, Akamatsu Y. Effect of CpG methylation on RAG1/RAG2 reactivity: implications of direct and indirect mechanisms for controlling V(D)J cleavage. EMBO Rep 2003;4(8): 774–80.
[160] Victora GD, Nussenzweig MC. Germinal centers. Annu Rev Immunol 2012;30:429–57.
[161] Meyer-Hermann M, Mohr E, Pelletier N, Zhang Y, Victora GD, Toellner KM. A theory of germinal center B cell selection, division, and exit. Cell Rep 2012;2(1):162–74.
[162] Dent AL, Shaffer AL, Yu X, Allman D, Staudt LM. Control of inflammation, cytokine expression andgerminal center formation by BCL-6. Science 1997;276:589–92.
[163] Ye BH, Cattoretti G, Shen Q, et al. The BCL-6 proto-oncogene controlsgerminal-centre formation and Th2-type inflammation. Nat Genet 1997;16:161–70.
[164] Fujita N, Jaye DL, Geigerman C, Akyildiz A, Mooney MR, Boss JM, Wade PA. MTA3 and theMi-2/NuRD complex regulate cell fate during B lymphocyte differentiation. Cell 2004;119:75–86.
[165] Fujita N, Jaye DL, Kajita M, Geigerman C, Moreno CS, Wade PA. MTA3, a Mi-2/NuRD complex subunit, regulates an invasive growth pathway in breast cancer. Cell 2003;113:207–19.
[166] Wu H, Deng Y, Feng Y, Long D, Ma K, Wang X, Zhao M, Lu L, Lu Q. Epigenetic regulation in B-cell maturation and its dysregulation in autoimmunity. Cell Mol Immunol 2018;15:676–84.
[167] Su ST, Ying HY, Chiu YK, Lin FR, Chen MY, Lin KI. Involvement of histone demethylase LSD1 in Blimp-1-mediated gene repression during plasma cell differentiation. Mol Cell Biol 2009;29: 1421–31.
[168] Shapiro-Shelef M, Calame K. Regulation of plasma-cell development. Nat Rev Immunol 2005;5:230–42.
[169] Yu J, Angelin-Duclos C, Greenwood J, Liao J, Calame K. Transcriptional repression by blimp-1 (PRDI-BF1) involves recruitment of histone deacetylase. Mol Cell Biol 2000;20:2592–603.
[170] Bell JJ, Bhandoola A. The earliest thymic progenitors for T cells possess myeloid lineage potential. Nature 2008;452:764–7.
[171] Wada H, Masuda K, Satoh R, Kakugawa K, Ikawa T, Katsura Y, Kawamoto H. Adult T-cell progenitors retain myeloid potential. Nature 2008;452:768–72.
[172] Rothenberg EV. Transcriptional control of early T and B cell developmental choices. Annu Rev Immunol 2014;32:283–321.
[173] Yui MA, Feng N, Rothenberg EV. Fine-scale staging of T cell lineage commitment in adult mouse thymus. J Immunol 2010;185:284–93.
[174] Masuda K, Kakugawa K, Nakayama T, Minato N, Katsura Y, Kawamoto H. T cell lineage determination precedes the initiation of TCR beta gene rearrangement. J Immunol 2007;179:3699–706.
[175] Krangel MS. Mechanics of T cell receptor gene rearrangement. Curr Opin Immunol 2009;21:133–9.
[176] Koltsova EK, Ciofani M, Benezra R, Miyazaki T, Clipstone N, Zúñiga-Pflücker JC, Wiest DL. Early growth response 1 and NF-ATc1 act in concert to promote thymocyte development beyond the beta–selection checkpoint. J Immunol 2007;179:4694–703.

[177] Germain RN. T-cell development and the CD4-CD8 lineage decision. Nat Rev Immunol 2002;2:309–22.

[178] Starr TK, Jameson SC, Hogquist KA. Positive and negative selection of T cells. Annu Rev Immunol 2003;21:139–76.

[179] Carpenter AC, Bosselut R. Decision checkpoints in the thymus. Nat Immunol 2010;11:666–73.

[180] Swat W, Ignatowicz L, von Boehmer H, Kisielow P. Clonal deletion of immature CD4+8+ thymocytes in suspension culture by extrathymic antigenpresenting cells. Nature 1991;351:150–3.

[181] Vasquez NJ, Kaye J, Hedrick SM. In vivo and in vitro clonal deletion of double-positive thymocytes. J Exp Med 1992;175:1307–16.

[182] Murphy KM, Heimberger AB, Loh DY. Induction by antigen of intrathymic apoptosis of CD4+CD8+TCRlo thymocytes in vivo. Science 1990;250:1720–3.

[183] Kappler JW, Roehm N, Marrack P. T cell tolerance by clonal elimination in the thymus. Cell 1987;49:273–80.

[184] Palmer E. Negative selection–clearing out the bad apples from the T-cell repertoire. Nat Rev Immunol 2003;3:383–91.

[185] Singer A, Adoro S, Park JH. Lineage fate and intense debate: myths, models and mechanisms of CD4- versus CD8-lineage choice. Nat Rev Immunol 2008;8:788–801.

[186] Janeway Jr. CA. The T cell receptor as a multicomponent signalling machine: CD4/CD8 coreceptors and CD45 in T cell activation. Annu Rev lmmunol 1992;10:645–74.

[187] Li L, Leid M, Rothenberg EV. An early T cell lineage commitment checkpoint dependent on the transcription factor Bcl11b. Science 2010;329(5987):89–93.

[188] Li P, Burke S, Wang J, Chen X, Ortiz M, Lee SC, Lu D, Campos L, Goulding D, Ng BL, Dougan G, Huntly B, Gottgens B, Jenkins NA, Copeland NG, Colucci F, Liu P. Reprogramming of T cells to natural killer-like cells upon Bcl11b deletion. Science 2010;329(5987):85–9.

[189] Ikawa T, Hirose S, Masuda K, Kakugawa K, Satoh R, Shibano-Satoh A, Kominami R, Katsura Y, Kawamoto H. An essential developmental checkpoint for production of the T cell lineage. Science 2010;329(5987):93–6.

[190] Hu G, Cui K, Fang D, Hirose S, Wang X, Wangsa D, Jin W, Ried T, Liu P, Zhu J, Rothenberg EV, Zhao K. Transformation of accessible chromatin and 3D nucleome underlies lineage commitment of early T cells. Immunity 2018;48 [227-42.e8].

[191] Ting CN, Olson MC, Barton KP, Leiden JM. Transcription factor GATA-3 is required for development of the T-cell lineage. Nature 1996;384(6608):474–8.

[192] Weber BN, Chi AW, Chavez A, Yashiro-Ohtani Y, Yang Q, Shestova O, Bhandoola A. A critical role for TCF-1 in T-lineage specification and differentiation. Nature 2011;476:63–8.

[193] Germar K, Dose M, Konstantinou T, Zhang J, Wang H, Lobry C, Arnett KL, Blacklow SC, Aifantis I, Aster JC, Gounari F. T-cell factor 1 is a gatekeeper for T-cell specification in response to notch signaling. Proc Natl Acad Sci U S A 2011;108:20060–5.

[194] Emmanuel AO, Arnovitz S, Haghi L, Mathur PS, Mondal S, Quandt J, Okoreeh MK, Maienschein-Cline M, Khazaie K, Dose M, Gounari F. TCF-1 and HEB cooperate to establish the epigenetic and transcription profiles of CD4+CD8+ thymocytes. Nat Immunol 2018;19:1366–78.

[195] Xing S, Li F, Zeng Z, Zhao Y, Yu S, Shan Q, Li Y, Phillips FC, Maina PK, Qi HH, Liu C, Zhu J, Pope RM, Musselman CA, Zeng C, Peng W, Xue H-H. Tcf1 and Lef1 transcription factors establish CD8+ T cell identity through intrinsic HDAC activity. Nat Immunol 2016;17:695–703.

[196] Steinke FC, Yu S, Zhou X, He B, Yang W, Zhou B, Kawamoto H, Zhu J, Tan K, Xue H-H. TCF-1 and LEF-1 act upstream of Th-POK to promote the CD4(+) T cell fate and interact with Runx3 to silence Cd4 in CD8(+) T cells. Nat Immunol 2014;15:646–56.

[197] Johnson JL, Georgakilas G, Petrovic J, Kurachi M, Cai S, Harly C, Pear WS, Bhandoola A, Wherry EJ, Vahedi G. Lineage determining transcription factor TCF-1 initiates the epigenetic identity of T cells. Immunity 2018;48 [243-57.e10].

[198] Rodriguez RM, Suarez-Alvarez B, Mosén-Ansorena D, García-Peydró M, Fuentes P, García-León MJ, Gonzalez-Lahera A, Macias-Camara N, Toribio ML, Aransay AM, Lopez-Larrea C. Regulation of the transcriptional program by DNA methylation during human alpha/beta T-cell development. Nucleic Acids Res 2015;43(2):760–74.

[199] Ji H, Ehrlich LI, Seita J, Murakami P, Doi A, Lindau P, Lee H, Aryee MJ, Irizarry RA, Kim K, Rossi DJ, Inlay MA, Serwold T, Karsunky H, Ho L, Daley GQ, Weissman IL, Feinberg AP. Comprehensive methylome map of lineage commitment from haematopoietic progenitors. Nature 2010;467(7313):338–42.

[200] Lee PP, Fitzpatrick DR, Beard C, Jessup HK, Lehar S, Makar KW, Pérez-Melgosa M, Sweetser MT, Schlissel MS, Nguyen S, Cherry SR, Tsai JH, Tucker SM, Weaver WM, Kelso A, Jaenisch R, Wilson CB. A critical role for Dnmt1 and DNA methylation in T cell development, function, and survival. Immunity 2001;15(5):763–74.

[201] He X, He X, Dave VP, Zhang Y, Hua X, Nicolas E, Xu W, Roe BA, Kappes DJ. The zinc finger transcription factor Th-POK regulates CD4 versus CD8 T-cell lineage commitment. Nature 2005; 433(7028):826–33.

[202] Kakugawa K, Kojo S, Tanaka H, Seo W, Endo TA, Kitagawa Y, Muroi S, Tenno M, Yasmin N, Kohwi Y, Sakaguchi S, Kowhi-Shigematsu T, Taniuchi I. Essential roles of SATB1 in specifying T lymphocyte subsets. Cell Rep 2017;19(6):1176–88.

[203] Kondo M, Tanaka Y, Kuwabara T, Naito T, Kohwi-Shigematsu T, Watanabe A. SATB1 plays a critical role in establishment of immune tolerance. J Immunol 2016;196(2):563–72.

[204] Alvarez JD, Yasui DH, Niida H, Joh T, Loh DY, Kohwi-Shigematsu T. The MAR-binding protein SATB1 orchestrates temporal and spatial expression of multiple genes during T-cell development. Gene Dev 2000;14(5):521–35.

[205] Wildt KF, Sun G, Grueter B, Fischer M, Zamisch M, Ehlers M, Bosselut R. The transcription factor Zbtb7b promotes CD4 expression by antagonizing Runx-mediated activation of the CD4 silencer. J Immunol 2007;179(7):4405–14.

[206] Woolf E, Xiao C, Fainaru O, Lotem J, Rosen D, Negreanu V, et al. Runx3 and Runx1 are required for CD8 T cell development during thymopoiesis. Proc Natl Acad Sci U S A 2003; 100(13):7731–6.

[207] Egawa T, Littman DR. ThPOK acts late in specification of the helper T cell lineage and suppresses Runx-mediatedcommitment to the cytotoxic T cell lineage. Nat Immunol 2008;9(10):1131–9.

[208] Fontenot JD, Gavin MA, Rudensky AY. Foxp3 programs the development and function of CD41 CD251 regulatory T cells. Nat Immunol 2003;4:330–6.

[209] Hori S, Nomura T, Sakaguchi S. Control of regulatory T cell development by the transcription factor Foxp3. Science 2003;299:1057–61.

[210] Lio CW, Hsieh CS. A two-step process for thymic regulatory T cell development. Immunity 2008;28:100–11.

[211] Kitagawa Y, Ohkura N, Kidani Y, Vandenbon A, Hirota K, Kawakami R, Yasuda K, Motooka D, Nakamura S, Kondo M, Taniuchi I, Kohwi-Shigematsu T, Sakaguchi S. Guidance of regulatory T cell development by Satb1-dependent super-enhancer establishment. Nat Immunol 2017;18:173–83.

[212] Ohkura N, Hamaguchi M, Morikawa H, Sugimura K, Tanaka A, Ito Y, Osaki M, Tanaka Y, Yamashita R, Nakano N, Huehn J, Fehling HJ, Sparwasser T, Nakai K, Sakaguchi S. T cell receptor stimulation-induced epigenetic changes and Foxp3 expression are independent and complementary events required for Treg cell development. Immunity 2012;37:785–99.

[213] Bevington SL, Cauchy P, Piper J, Bertrand E, Lalli N, Jarvis RC, Gilding LN, Ott S, Bonifer C, Cockerill PN. Inducible chromatin priming is associated with the establishment of immunological memory in T cells. EMBO J 2016;35:515–35.

[214] Allison KA, Sajti E, Collier JG, Gosselin D, Troutman TD, Stone EL, Hedrick SM, Glass CK. Affinity and dose of TCR engagement yield proportional enhancer and gene activity in CD41 T cells. Elife 2016;5: e10134.

[215] Mosmann TR, Coffman RL. TH1 and TH2 cells: different patterns of lymphokine secretion lead to different functional properties. Annu Rev Immunol 1989;7:145–73.

[216] Notani D, Gottimukkala KP, Jayani RS, Limaye AS, Damle MV, Mehta S, Purbey PK, Joseph J, Galande S. Global regulator SATB1 recruits beta-catenin and regulates TH2 differentiation in Wnt-dependent manner. PLoS Biol 2010;8:e1000296.

[217] Schmidl C, Delacher M, Huehn J, Feuerer M. Epigenetic mechanisms regulating T-cell responses. J Allergy Clin Immunol 2018;142:3.

[218] Wei G, Wei L, Zhu J, Zang C, Hu-Li J, Yao Z, Cui K, Kanno Y, Roh T-Y, Watford WT, Schones DE, Peng W, Sun H-w, Paul WE, O'Shea JJ, Zhao K. Global mapping of H3K4me3 and H3K27me3 reveals specificity and plasticity in lineage fate determination of differentiating CD4+ T cells. Immunity 2009;30(1):155–67.

[219] Sekiya T. Immune cell development and epigenetics. In: The epigenetics of autoimmunity. Elsevier Inc; 2019. https://doi.org/10.1016/B978-0-12-809912-4.00002-7.

[220] Schoenborn JR, Dorschner MO, Sekimata M, Santer DM, Shnyreva M, Fitzpatrick DR, Stamatoyannopoulos JA, Wilson CB. Comprehensive epigenetic profiling identifies multiple distal regulatory elements directing transcription of the gene encoding interferongamma. Nat Immunol 2007;8(7):732–42.

[221] Santangelo S, Cousins DJ, Winkelmann N, Triantaphyllopoulos K, Staynov DZ. Chromatin structure and DNA methylation of the IL-4 gene in human T(H)2 cells. Chromosome Res 2009;17(4):485–96.

[222] Yue X, Trifari S, Äijö T, Tsagaratou A, Pastor WA, Zepeda-Martínez JA, Lio CW, Li X, Huang Y, Vijayanand P, Lähdesmäki H, Rao A. Control of Foxp3 stability through modulation of TET activity. J Exp Med 2016;213(3):377–97.

[223] Toker A, Engelbert D, Garg G, Polansky JK, Floess S, Miyao T, Baron U, Düber S, Geffers R, Giehr P, Schallenberg S, Kretschmer K, Olek S, Walter J, Weiss S, Hori S, Hamann A, Huehn J. Active demethylation of the Foxp3 locus leads to the generation of stable regulatory T cells within the thymus. J Immunol 2013;190(7):3180–8.

[224] Cain CE, Blekhman R, Marioni JC, Gilad Y. Gene expression differences among primates are associated with changes in a histone epigenetic modification. Genetics 2011;187(4):1225–34.

[225] Bernstein BE, Meissner A, Lander ES. The mammalian epigenome. Cell 2007;128(4):669–81.

[226] Lieberman-Aiden E, van Berkum NL, Williams L, Imakaev M, Ragoczy T, Telling A, Amit I, Lajoie BR, Sabo PJ, Dorschner MO, Sandstrom R, Bernstein B, Bender MA, Groudine M, Gnirke A, Stamatoyannopoulos J, Mirny LA, Lander ES, Dekker J. Comprehensive mapping of long-range interactions reveals folding principles of the human genome. Science 2009;326:289–93.

[227] Boucheron N, Tschismarov R, Goeschl L, Moser MA, Lagger S, Sakaguchi S, Winter M, Lenz F, Vitko D, Breitwieser FP, Müller L, Hassan H, Bennett KL, Colinge J, Schreiner W, Egawa T, Taniuchi I, Matthias P, Seiser C, Ellmeier W. CD4(1) T cell lineage integrity is controlled by the histone deacetylases HDAC1 and HDAC2. Nat Immunol 2014;15:439–48.

[228] Durek P, Nordstrom K, Gasparoni G, Salhab A, Kressler C, de Almeida M, et al. Epigenomic profiling of human CD4(1) T cells supports a linear differentiation model and highlights molecular regulators of memory development. Immunity 2016;45:1148–61.

[229] Barski A, Cuddapah S, Kartashov AV, Liu C, Imamichi H, Yang W, Peng W, Lane HC, Zhao K. Rapid recall ability of memory T cells is encoded in their epigenome. Sci Rep 2017;7:39785.

[230] Zediak VP, Johnnidis JB, Wherry EJ, Berger SL. Cutting edge: persistently open chromatin at effector gene loci in resting memory CD8+ T cells independent of transcriptional status. J Immunol 2011;186(5):2705–9.

[231] Feng S, Jacobsen SE, Reik W. Epigenetic reprogramming in plant and animal development. Science 2010;330(6004):622–7.

[232] Zemach A, McDaniel IE, Silva P, Zilberman D. Genome-wide evolutionary analysis of eukaryotic DNA methylation. Science 2010;328(5980):916–9.

[233] Maunakea AK, Nagarajan RP, Bilenky M, Ballinger TJ, D'Souza C, Fouse SD, Johnson BE, Hong C, Nielsen C, Zhao Y, Turecki G, Delaney A, Varhol R, Thiessen N, Shchors K, Heine VM, Rowitch DH, Xing X, Fiore C, Schillebeeckx M, Jones SJ, Haussler D, Marra MA, Hirst M, Wang T, Costello JF. Conserved role of intragenic DNA methylation in regulating alternative promoters. Nature 2010;466(7303):253–7.

[234] Xiao S, Xie D, Cao X, Yu P, Xing X, Chen CC, Musselman M, Xie M, West FD, Lewin HA, Wang T, Zhong S. Comparative epigenomic annotation of regulatory DNA. Cell 2012;149(6):1381–92.

[235] Long HK, Sims D, Heger A, Blackledge NP, Kutter C, Wright ML, Grützner F, Odom DT, Patient R, Ponting CP, Klose RJ. Epigenetic conservation at gene regulatory elements revealed by non-methylated DNA profiling in seven vertebrates. Elife 2013;2:e00348.

[236] Pai AA, Bell JT, Marioni JC, Pritchard JK, Gilad Y. A genome-wide study of DNA methylation patterns and gene expression levels in multiple human and chimpanzee tissues. PLoS Genet 2011;7(2): e1001316.

[237] Zhou J, Sears RL, Xing X, Zhang B, Li D, Rockweiler NB, Jang HS, Choudhary MNK, Lee HJ, Lowdon RF, Arand J, Tabers B, Gu CC, Cicero TJ, Wang T. Tissue-specific DNA methylation is conserved across human, mouse, and rat, and driven by primary sequence conservation. BMC Genomics 2017;18(1):724.

[238] Shnyreva M, Weaver WM, Blanchette M, Taylor SL, Tompa M, Fitzpatrick DR, Wilson CB. Evolutionarily conserved sequence elements that positively regulate IFN-gamma expression in T cells. Proc Natl Acad Sci U S A 2004;101(34):12622–7.

[239] Avni O, Lee D, Macian F, Szabo SJ, Glimcher LH, Rao A. T(H) cell differentiation is accompanied by dynamic changes in histone acetylation of cytokine genes. Nat Immunol 2002;3(7):643–51.

[240] Fields PE, Kim ST, Flavell RA. Cutting edge: changes in histone acetylation at the IL-4 and IFN-gamma loci accompany Th1/Th2 differentiation. J Immunol 2002;169(2):647–50.

[241] Zhou W, Chang S, Aune TM. Long-range histone acetylation of the Ifng gene is an essential feature of T cell differentiation. Proc Natl Acad Sci U S A 2004;101(8):2440–5.

[242] Balasubramani A, Winstead CJ, Turner H, Janowski KM, Harbour SN, Shibata Y, Crawford GE, Hatton RD, Weaver CT. Deletion of a conserved cis-element in the Ifng locus highlights the role of acute histone acetylation in modulating inducible gene transcription. PLoS Genet 2014;10(1) e1003969.

[243] Loots GG, Locksley RM, Blankespoor CM, Wang ZE, Miller W, Rubin EM, Frazer KA. Identification of a coordinate regulator of interleukins 4, 13, and 5 by cross-species sequence comparisons. Science 2000;288(5463):136–40.

[244] Spilianakis CG, Flavell RA. Long-range intrachromosomal interactions in the T helper type 2 cytokine locus. Nat Immunol 2004;5(10):1017–27.

[245] Lee GR, Spilianakis CG, Flavell RA. Hypersensitive site 7 of the TH2 locus control region is essential for expressing TH2 cytokine genes and for long-range intrachromosomal interactions. Nat Immunol 2004;6(1):42–8.

[246] Lee DU, Rao A. Molecular analysis of a locus control region in the T helper 2 cytokine gene cluster: a target for STAT6 but not GATA3. Proc Natl Acad Sci U S A 2004;101(45):16010–5.

[247] Fields PE, Lee GR, Kim ST, Bartsevich VV, Flavell RA. Th2-specific chromatin remodeling and enhancer activity in the Th2 cytokine locus control region. Immunity 2004;21(6):865–76.

[248] Lee GR, Fields PE, Griffin TJ, Flavell RA. Regulation of the Th2 cytokine locus by a locus control region. Immunity 2003;19(1):145–53.

[249] Yasui D, Miyano M, Cai S, Varga-Weisz P, Kohwi-Shigematsu T. SATB1targets chromatin remodelling to regulate genes over long distances. Nature 2002;419:641–5.

[250] Cai S, Lee CC, Kohwi-Shigematsu T. SATB1 packages densely looped, transcriptionally active chromatin for coordinated expression of cytokine genes. Nat Genet 2006;38(11):1278–88.

[251] Kumar PP, Purbey PK, Sinha CK, Notani D, Limaye A, Jayani RS, Galande S. Phosphorylation of SATB1, a global gene regulator, acts as a molecular switch regulating its transcriptional activity in vivo. Mol Cell 2006;22:231–43.

[252] Notani D, Gottimukkala KP, Jayani RS, Limaye A, Damle MV, Mehta SM, Purbey PK, Joseph J, Galande S. Global regulator SATB1 recruits β-catenin and regulates T_H2 differentiation in Wnt-dependent manner. PLoS Biol 2010;8(1): e1000296.

[253] Purbey PK, Singh S, Notani D, Kumar PP, Limaye AS, Galande S. Acetylation-dependent interaction of SATB1 and CtBP1 mediates transcriptional repression by SATB1. Mol Cell Biol 2009;29:1321–37.

Further reading

Verbeek S, Izon D, Hofhuis F, Robanus-Maandag E, te Riele H, van de Wetering M, Oosterwegel M, Wilson A, MacDonald HR, Clevers H. An HMG-box-containing T-cell factor required for thymocyte differentiation. Nature 1995;374(6517):70–4.

Vahedi G, Takahashi H, Nakayamada S, Sun HW, Sartorelli V, Kanno Y, et al. STATs shape the active enhancer landscape of T cell populations. Cell 2012;151:981–93.

Juelich T, Sutcliffe EL, Denton A, He Y, Doherty PC, Parish CR, Turner SJ, Tremethick DJ, Rao S. Interplay between chromatin remodeling and epigenetic changes during lineage-specific commitment to granzyme B expression. J Immunol 2009;183(11):7063–72.

Scharer CD, Barwick BG, Youngblood BA, Ahmed R, Boss JM. Global DNA methylation remodeling accompanies CD8 T cell effector function. J Immunol 2013;191(6):3419–29.

Kumar PP, Purbey PK, Ravi DS, Mitra D, Galande S. Displacement of SATB1-bound HDAC1 corepressor by HIV-1 transactivator induces expression of Interleukin-2 and its receptor in T cells. Mol Cell Biol 2005;25:1620–33.

CHAPTER 4

Epigenetic mechanisms in the regulation of lymphocyte differentiation

Nina Schmolka[a], **Bruno Silva-Santos**[b], **Anita Q. Gomes**[b,c]
[a]Department of Molecular Mechanisms of Disease, University of Zurich, Zurich, Switzerland
[b]Instituto de Medicina Molecular João Lobo Antunes, Faculdade de Medicina, Universidade de Lisboa, Lisboa, Portugal
[c]H&TRC Health & Technology Research Center, ESTeSL—Escola Superior de Tecnologia da Saúde, Instituto Politécnico de Lisboa, Lisbon, Portugal

Contents

Introduction 77
Part I: Chromatin-based epigenetic mechanisms in lymphocyte differentiation 80
    DNA methylation and lineage commitment 80
    DNA methylation writers, readers, and erasers in immune cell differentiation 82
    Histone modifications in lymphocyte differentiation 84
    Histone modification writers, readers, and erasers in lymphocyte differentiation 86
    Chromatin accessibility in the differentiation of immune cells 88
Part II: RNA-based mechanisms of regulation of lymphocyte differentiation 89
    MicroRNA-mediated regulation of T- and B-cell differentiation 90
    Specific miRNAs regulating T- and B-cell differentiation 92
    The emerging role of lncRNAs in immune cell differentiation 100
    LncRNAs in T-cell differentiation 101
    LncRNAs in B-cell differentiation 103
    Cross talk between noncoding RNAs 104
Concluding remarks 105
References 106
Further reading 116

Introduction

The immune system is an ideal model to investigate molecular mechanisms operating during cellular differentiation and establishment of cell fate and maintenance of phenotypic plasticity. Lymphocyte differentiation is one of the best-studied systems, as multiple developmental stages can be isolated; and the underlying transcriptional programs (including master and auxiliary factors) have been thoroughly dissected. There is an increasing understanding that the complex process of lymphocyte differentiation is controlled by the interplay of transcription factors, epigenetic mechanisms, and chromatin interactions. All such mechanisms are the foundation for the establishment and

Epigenetics of the Immune System
https://doi.org/10.1016/B978-0-12-817964-2.00004-6
© 2020 Elsevier Inc.
All rights reserved.

maintenance of distinct transcriptional patterns that ultimately allow differentiation and the appearance of distinct cell identities.

While the so-called "pioneering" factors and transcription factors are crucial to determine which lineage-specifying genes are activated or repressed, epigenetic mechanisms are also critical to regulating the established gene programs. Epigenetic mechanisms, which control gene expression without changing the underlying DNA sequence, include remodeling of the chromatin structure, DNA (de)methylation and histone modifications, as well as noncoding RNAs including microRNAs (miRNAs) and long-noncoding RNAs (lncRNAs). Over the last couple of decades, the emergence of high-throughput technologies has generated a tremendous amount of information regarding the genome-wide distribution of epigenetic modifiers and modifications in immune cells at different developmental stages, therefore shedding light on epigenetic changes occurring during cell differentiation. Additional single-cell approaches to profile the (epi-)transcriptome are arising, and we are entering a new era with the privilege to shift the main focus from generating large-scale data to integrating them. In this chapter, we discuss key epigenetic mechanisms controlling lymphocyte differentiation, from earlier stages (divergence from the myeloid branch) to the late acquisition of effector functions. We aim to elucidate key principles that have emerged from the investigation on writers, readers, and erasers of specific epigenetic modifications; studies focusing on the global characterization of the epigenetic landscape, and on the specific roles played by noncoding RNAs in the regulation of gene expression during lymphocyte differentiation.

T and B cells are two distinct cell lineages responsible for cellular and humoral immune responses, respectively. Both cell types develop from hematopoietic stem cells (HSC) that give rise to multipotent progenitors (MPPs) that further differentiate into lineage-restricted common lymphoid progenitors (CLP) [1] (Fig. 1). Subsequently, B-cell development takes place in the bone marrow by generating pro-B cells and further pre-B cells that start the recombination of the immunoglobulin chains. After successful recombination of the B-cell receptor (BCR) naïve immature B cells exit the bone marrow and differentiate after encountering their cognate antigen into mature B cells that form short-lived antibody-producing plasmablast, long-lived high-affinity antibody-producing plasma cells or memory B cells [2]. During an immune response, some activated B cells undergo class switch recombination (CSR) and affinity maturation of the BCR by somatic hypermutation (SHM) in the germinal centers of secondary lymphoid organs (Fig. 1).

On the other hand, T cell develops in the thymus where after successful V(D)J recombination of TCR loci, αβ, and γδ T cells are generated. αβ T cells further undergo positive selection into one of two subsets, CD4+ and CD8+ T cells, that respectively account for helper and cytotoxic functions in the periphery. There, and depending on the cytokine milieu produced during an immune response, naive CD4+ T cells differentiate into an array of functionally different effector T-helper

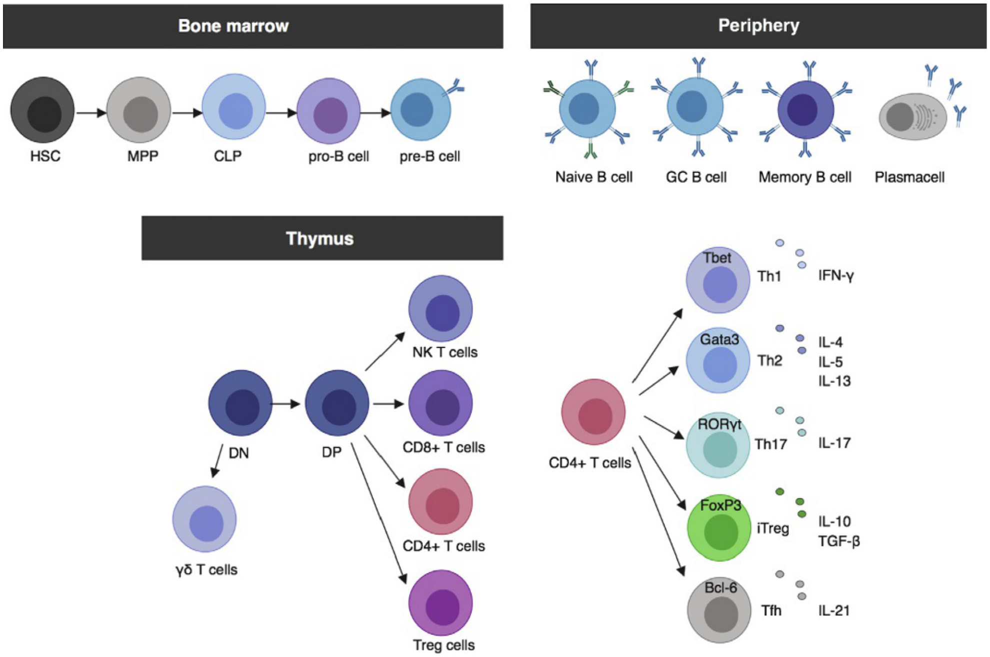

Bone marrow
HSC
MPP
CLP
pro-B cell
pre-B cell
Periphery
Naive B cell
GC B cell
Memory B cell
Plasmacell
Thymus
DN
DP
γδ T cells
NK T cells
CD8+ T cells
CD4+ T cells
Treg cells
CD4+ T cells
Tbet
Th1
IFN-γ
Gata3
Th2
IL-4
IL-5
IL-13
RORγt
Th17
IL-17
FoxP3
iTreg
IL-10
TGF-β
Bcl-6
Tfh
IL-21

Fig. 1 Lymphocytes differentiation. B and T cells develop from hematopoietic stem cells (HSCs) in the bone marrow in a stepwise manner from multipotent progenitors (MPP) toward common lymphoid progenitors (CLP). B-cell development continues in the bone marrow toward pro- and pre-B cells. In the periphery, naïve B cells are activated after antigen recognition and differentiate toward antibody-secreting plasmacells, memory B cells, and germinal center (GC) B cells. T-cell development occurs in the thymus. Double negative (DN, CD4-CD8-) cells and double positive (DP, CD4+CD8+) T cells, which give rise to γδ T cells, CD4+, CD8+, NKT, and regulatory T cells (Treg). In the periphery, naïve CD4+ T cells differentiate after antigen encounter regarding the specific cytokine milieu to several effector T-cell subsets including T helper (Th) 1, Th2, Th17, Treg, and Tfh cells characterized by the expression of master transcription factors and secretion of signature cytokines. *(Figure created with Biorender.com.)*

(Th) subsets. Th-cell subsets are distinguished based on the master transcription factors that drive their differentiation and signature cytokine profiles [3–7]. Several Th subsets have been characterized, including Th1 (defined by IFN-γ production), Th2 (defined by IL-4, IL-5, and IL-13 production), Th17 (defined by IL-17 production) and Th9 (defined by IL-9 production). *Tbx1*/T-bet, Gata-3, *Rorc*/ROTγt, and IRF4 are the lineage-defining transcription factors (TFs), respectively [8]. Other crucial CD4+ T-cell subsets are regulatory T cells (Treg cells) that are essential for maintaining self-tolerance and immunological homeostasis. Foxp3 is the lineage-defining TF of Treg cells, which can be either generated in the thymus (tTreg cells) or in the periphery, i.e., induced upon activation under polarizing conditions (iTreg cells) [9] (Fig. 1).

Part I: Chromatin-based epigenetic mechanisms in lymphocyte differentiation

In this first part, we will discuss key epigenetic mechanisms implicated in the differentiation of diverse subsets of lymphoid cells. We also refer to the myriad of excellent reviews that focus more precisely on a particular differentiation step and/or the underlying epigenetic mechanisms [10–21].

DNA methylation and lineage commitment

DNA methylation of cytosine residues within CpG dinucleotides is a key epigenetic mark in development and cell differentiation mainly associated with gene silencing [22]. DNA methylation is mediated by DNA methyltransferases (DNMTs), whereas DNMT1 is mainly attributed to the maintenance of DNA methylation and DNMT3A and DNMT3B are responsible for setting de novo methylation [23, 24] (Fig. 2). Until

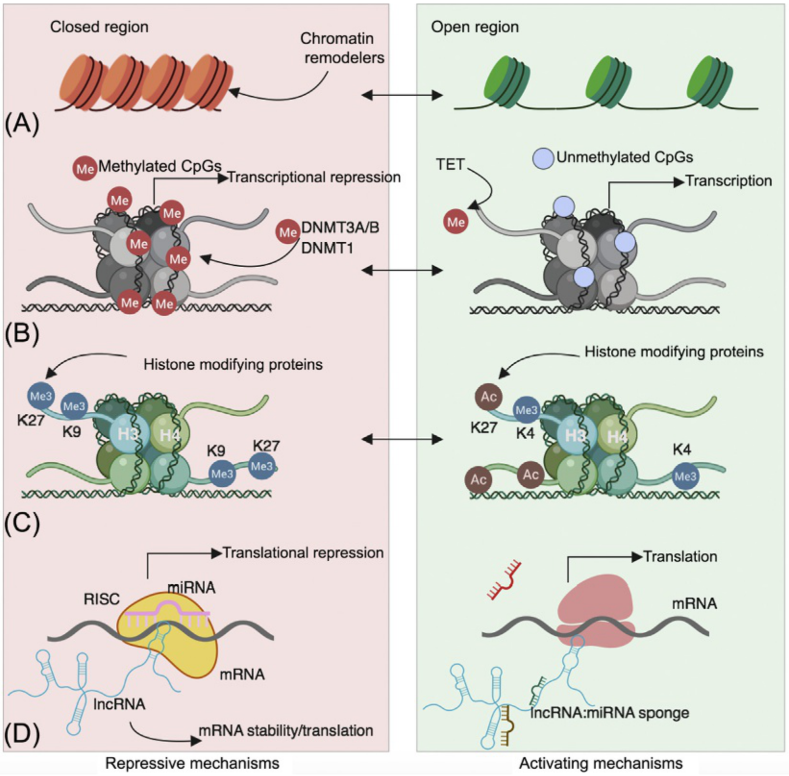

Closed region
Chromatin remodelers
Open region
(A)
Me Methylated CpGs
Transcriptional repression
Me DNMT3A/B DNMT1
TET
Unmethylated CpGs
Transcription
Me
(B)
Histone modifying proteins
Me3
K27
K9
H3
H4
Ac
K4
(C)
Translational repression
RISC
miRNA
mRNA
lncRNA
mRNA stability/translation
Translation
lncRNA:miRNA sponge
(D)
Repressive mechanisms
Activating mechanisms

Fig. 2 Chromatin- and noncoding RNA-based epigenetic mechanisms modify gene expression. Specific alterations in chromatin including chromatin accessibility (A), DNA methylation (B), histone tail modifications (C), and expression of noncoding RNAs (microRNAs (miRNAs) and long noncoding RNAs (lncRNAs) (D) are mechanisms that contribute to gene regulation which ultimately establish a repressive or active state. DNA methyltransferases (DNMT), Ten-eleven translocation enzymes (TET). *(Figure created with Biorender.com.)*

today DNA methylation is the bona fide epigenetic modification that can faithfully propagate its information from one cell generation to the next. This heritability is mainly achieved through the active mechanism of DNA methylation maintenance by DNMT1 that copies the pattern of methylated cytosines from parental to progeny DNA strands [21, 25]. By silencing specific gene programs, DNA methylation is a key epigenetic modification that influences cell type- and context-specific gene expression, ultimately impacting on cell differentiation [26]. Functionally DNA methylation at gene promoters and distal regulatory elements can mediate gene repression by attraction or repulsion of regulatory factors, e.g., directly by blocking the binding of methylation-sensitive transcription factors and/or indirectly by recruiting chromatin modifiers to methylated cytosine which either set other silencing epigenetic marks or mediate chromatin compaction [27–30]. DNA demethylation processes can be achieved either passively through the absence of maintenance DNA methylation which leads to loss of DNA methylation during replication or actively by a newly discovered family of proteins, the Ten-Eleven Translocation (TET) proteins which convert 5-methylcytosine (5mC) to its oxidized derivative 5-hydroxymethylcytosine (5hmC) [31–33], which can be removed by DNA repair mechanisms.

The first genome-wide characterization of DNA methylation dynamics during blood cells development used gene array technology in several hematopoietic progenitors [34]. This study revealed that early stages in blood cell lineage commitment show a distinct degree of total DNA methylation and methylation changes were detected from MPPs toward CLPs, common myeloid progenitors (CMPs), and thymocyte progenitors [34]. Importantly, regulatory factors like Meis1, associated with locking cells in an undifferentiated state became methylated and silenced in the course of the differentiation process [35]. Vice versa lineages determining genes became demethylated in the respective lineage, e.g., Lck, an Src family kinase member responsible for initiating signaling downstream of the T-cell receptor (TCR) [36] shows demethylation in T-cell precursors [34]. Interestingly, lymphoid cells, in general, display an overall increased level of total DNA methylation compared to myeloid compartment that shows a hypomethylated pattern similar to that found in progenitor populations [34]. In agreement with the methylation changes detected in early hematopoiesis, several studies addressed the DNA methylation pattern in terminally differentiated immune cell subsets including T cells, B cells, NK cells, monocytes, and other innate immune cells. It is well understood that distinct subsets show a lineage-specific DNA methylation signature that can be used to distinguish them [37]. Furthermore, a good model to study the impact of DNA methylation is the process of CD4+ helper T-cell differentiation. Analysis of three major Th subsets, Th1, Th2, and Th17 showed lineage-specific demethylation patterns in the corresponding signature cytokine loci. *Ifng* hypermethylated in naïve CD4+ T cells shows hypomethylation in Th1 cells but remains methylated in Th2 cells [38], IL-4 promoter regions are demethylated in Th2 cells [39], and the IL-17 locus is demethylated in Th17 cells [40].

DNA methylation writers, readers, and erasers in immune cell differentiation

To get a better insight in the functional role of DNA methylation besides monitoring global DNA methylation dynamics in distinct immune cell subsets, several studies address the role of key enzymes setting, reading, or removing DNA methylation. Studies on DNMTs can tackle question regarding the impact of demethylation or de novo methylation events during immune cell differentiation. As DNMT1 KO mice are not viable, conditional DNMT1 knockout or DNMT1-hypomorph mice were used. A block in DNA methylation maintenance decreased HSC self-renewal potential and led to defective differentiation pattern toward increased myeloid-erythroid lineage commitment and a defect in lymphoid lineage differentiation [41, 42]. This goes along with the observation that myeloid cells are hypomethylated and resemble more closely the methylation pattern of HSCs opposed to a more hypermethylated state in lymphoid cells [34, 43]. Therefore the premature demethylation of specific genes in DNMT1-deficient HSCs prevents the suppression of key myeloerythroid regulators thus giving rise to myeloerythroid but not lymphoid progeny [41, 42]. On the other hand, commitment to the lymphoid lineage depends on the increase of DNA methylation level but, interestingly, constitutive methylation is not continuously required throughout lymphopoiesis [41, 42]. If DNMT1-deficiency is reduced after commitment to B-cell lineage (using CD19-cre mice), B-cell identity and maturation are maintained, pointing to a different DNA methylation dependency in lineage commitment and maintenance [41]. Additionally, the importance of DNA methylation in terminal effector differentiation is further underlined by early studies where naïve CD4+ and CD8+ T cells with conditional deletion of *Dnmt1* or treated with a demethylating agent (5-Aza-2-deoxycytidine) exhibit increased frequency of effector cells expressing cytokines, including, IL-2, IFN- γ, and IL-4 [44, 45].

Surprisingly, the lack of both de novo methyltransferases DNMT3A and DNMT3B did not impact on HSC differentiation as all lineages were generated in a normal ratio but DNMT3A/3B double deficiency mainly decreased HSCs self-renewal potential [46]. In agreement with this study, it was recently shown that genes with a crucial function in hematopoietic progenitors are downregulated during T-cell commitment without showing concurrent DNA methylation changes [47]. That lymphoid commitment occurs without setting de novo methylation points to additional mechanisms that reinforce silencing of gene expression independent of DNA methylation such as repressive marks on histones.

Research on T cells pointed to the role of DNMT3a in maintaining lineage stability of terminally differentiated T cells [40, 48]. Both studies highlight the role of DNMT3a in setting de novo methylation at signature cytokine loci to restrict the expression of *Ifng* in Th2 cells and *Il13* in an allergic asthma model [40, 48]. Additionally, a T cell-specific KO of DNMT3a led to a bias of CD8 effector and memory cell differentiation [49]. Dnmt3a KO resulted in an increase of CD8+ memory precursor effector cells and long-term T-cell memory in three different acute viral infection models [49]. As a mechanism, the authors claimed that de novo methylation is necessary to methylate and silence TCF1, a crucial TF in T cells, whose expression drops in effector CD8+ T cells compared to naïve CD8+ T cells, but gets reexpressed in memory T cells [49]. Collectively, de novo methylation seems to be dispensable for the development of the major blood lineage but is crucial to fine-tune the differentiation toward specific effector lineages.

Besides DNA methyltransferases, DNA-methylation readers influence immune cell differentiation. A T cell-specific knockout of MeCP2, a nuclear protein that binds to cytosine-methylated DNA within dinucleotide CpG elements [50, 51] led to a block in naïve T-cell differentiation toward Th1 and Th17 T cells. Mechanistically MeCP2 is needed for the expression of a particular miRNA (miR-124) and MeCP2-deficient cells show decreased chromatin accessibility and decreased transcription of the miR-124 locus [52]. Ultimately, decreased miR-124 expression led to increased SOCS5 levels that inhibited cytokine-dependent activation of signal transducer and activator of transcription 1 (STAT1) and STAT3, which are necessary for the differentiation of Th1 and Th17 cells, respectively [52]. Another study revealed a function of MBD2, another methyl CpG-binding domain protein, in T-cell effector subset differentiation [53]. A loss of MBD2 led to enhanced expression of IL-4 in naïve T cells, as the binding of MBD2 at *Il4* prevents recruitment of the transcription factor GATA3 to this site that would induce IL-4 expression. Additionally, MBD2-deficient mice showed a skewed expression of IFN-γ in Th2 cells and Th2-type cytokines in Th1 cells and therefore MBD2 helps to restrict cytokine expression to the appropriate lineage [53].

Testing if DNA demethylation is necessary for the induction of a specific cell fate is experimentally difficult to tackle. Recent studies in T cells characterized the genome-wide patterns of 5hmC as an indirect readout for an active demethylation process [54, 55]. Dynamic changes of both 5mC and 5hmC were noted in many lineage-specific gene loci [54]. 5hmC marks were enriched in active and/or poised genes pointing to an active demethylation process during T-cell differentiation. Loss-of-function approaches where demethylation enzymes are deleted can further shed light on this question. Tet2 is the major Tet family member expressed in Th subsets. A recent study addressed the consequences of T cell-specific Tet2 deficiency on T-cell differentiation [54]. Loss of Tet2 impacted in the differentiation of specific T effector subsets. Whereas Th2 and iTregs showed normal 5hmC patterns (at signature genes), *Ifng*, *Il17*, and *Il10* showed significant decreased 5hmC patterns in Tet2-deficient cells compared to WT cells, pointing to a defect in DNA demethylation in these loci [54]. In agreement with this 5hmC pattern, Tet2-deficient T cells had a reduced differentiation potential toward Th1 and Th17 cells in vitro compared to wild type; and showed decreased expression of IFN-γ and IL-17, respectively [54]. Additionally, Tet2 controlled IL-10, IFN-γ, and IL-17 production in vivo in an autoimmune model, where Tet2-deficient mice were more resistant to experimental autoimmune encephalomyelitis (EAE) induction compared to control mice [54].

TET enzymes play additional roles during B-cell differentiation, where double deficiency in Tet2 and Tet3 led to a partial block of pro-B to pre-B-cell differentiation [56].

Collectively, there is now a compelling body of evidence that DNA methylation dynamics change in immune cell differentiation and that specific DNA methylation states are established at lineage-restricted cytokine and transcription factor genes, as well as at regulatory regions. Future challenges will be to decipher the causal role of DNA methylation in driving gene expression programs and how it is integrated into the rest of the epigenetic landscape. It is expected that the initial event to set site-specific methylation must involve recognition of particular sites in the genome by proteins like transcription factors which can recruit additional chromatin-modifying proteins that ultimately lead to an alteration in DNA methylation, thus conveying stable changes in gene structure that are maintained throughout time even in the absence of the initial factor [20]. It is still not known how the multistep establishment of site-specific de novo methylation or demethylation events is achieved. Important hints come from studies addressing the regulation of writers, readers, and erasers of DNA methylation and which cues in the epigenetic landscape influence them. For example, specific domains of DNMTs and cofactors interact with histone H3 tails and their recruitment is either blocked (by H3K4me3) [57–59] or attracted to sites with specific histone marking (e.g., H3K9me3 and H3K36me3) [60–64].

Histone modifications in lymphocyte differentiation

The basic structure of chromatin is the DNA, which is wrapped around nucleosomes. Nucleosomes are comprised of eight core histones proteins, pairs of H2A, H2B, H3, and H4 and approximately 147 base pairs of DNA is wound around them. One of the key epigenetic modifications is posttranslational marks on histone tails such as acetylation, methylation, ubiquitinylation, and sumoylation of lysine, methylation of arginine, and phosphorylation of serine. Importantly, the combination of these modifications contributes to the structure of chromatin and influence accessibility of the underlying DNA for transcriptional regulators that ultimately dictate gene transcription (Fig. 2). Additionally, the diverse set of histone modifications serves as a scaffold for other chromatin regulating factors, and therefore both changes in chromatin accessibility and recruitment of regulatory factors reinforce the activation or repression of gene expression. In the early 2000s, the concept of the histone code arose [65] to describe the cross talk between gene expression and histone modifications [66].

Over the last 25 years, there was an extensive amount of work characterizing histone modifications in cell differentiation and many of those seminal papers concentrated on immune cells. Some of the first studies on genome-wide distribution of a diverse set of histone marks were performed in human T cells [67, 68]. The initial studies were complemented with the characterization of histone modification in lymphoid development [69, 70]. By correlating the histone modification with gene expression profiles those studies are until today the basis of our current understanding of the histone landscape of active, poised, or repressed regions. Several modifications are associated with active gene expression. Among the best-studied examples is histone acetylation. There are 18 different histone lysine acetylation sites, which were all associated with active genes [68]. Genome-wide studies in lymphocytes showed a positive correlation of histone acetylation with actively transcribed genes both during CD4+ T-cell differentiation toward Th1 or Th2 cells and CD8+ T cells. Histone-H3 acetylation was correlated with active gene expression in cytokine genes and lineage-specifying genes in the studied T-cell subsets [71, 72]. Histone acetylation is thought to increase the accessibility of DNA to regulatory factors, by decreasing the positive charge of histones and therefore resulting in reduced affinity for negatively charged DNA [73]. Another important mark associated with active transcription is H3K4 methylation. H3K4me is a marker for active enhancers in combination with H3K27ac and p300 binding [74], whereas H3K4me2 and H3K36me3 are active marks mainly found throughout gene bodies and H4K4me3 at gene promoters [67, 68]. Interestingly, active promoters and enhancers are often depleted of nucleosomes or harbor unstable nucleosomes with histone variants like H3.3 [75, 76]. Those depleted regions lie upstream of the transcriptional start site and are around 150 bp and can be detected by DNase hypersensitivity assays. In T-cell differentiation those nucleosome-depleted sites have been mainly found in active cytokine loci [77].

Several histone marks are associated with silenced genes, the best-studied examples being H3K27me3 and H3K9me3. H3K27me3 associated with promoters and gene bodies of silenced genes and its distribution was extensively studied in CD4+ and CD8+ T-cell differentiation [72, 78–82]. An important finding through genome-wide characterization of histone marks in different T-cell subsets (including Th1, Th17, Th2, and Treg) was the identification of so-called bivalent or poised genes [79, 80]. Those bivalent genes show both active and repressive marks at the same loci, mainly H3K4me3 and H3K27me3, respectively. Bivalent marks are often found at key lineage-specifying genes, e.g., master transcription factors like *Tbx21* for Th1 and *Rorc* for Th17 lineage [80]. A poised state enables silenced genes to be rapidly induced under particular conditions. Therefore, if a cell differentiates toward its effector lineage, poised genes gain active histone marks and loserepressive marks (*Ifng* and *Tbx21* for Th1 cells, and *Il4*, *Il13*, and *Gata3* for Th2 cells) [80, 83].

The discovery of bivalent marking in T-cell subsets is also an indication of cellular plasticity. It is nowadays well understood that immune cells keep a form of functional plasticity that allows the differentiation toward a different lineage when exposed to strong repolarizing stimuli. For example, it was shown that Th17 cells [84, 85] and induced Treg (iTreg) [86, 87] can acquire the ability to produce IFN-γ both in vitro and in models of autoimmunity. Along the same lines, we showed that IL-17-producing γδ T cells can acquire IFN-γ expression under strong inflammatory conditions. Thus, by characterizing the H3K4me2 and H3K27me3 landscape in ex vivo isolated γδ T cells, we found that the γδ T-cell subset characterized by CD27 expression displayed a stable effector phenotype based on active marking on key lineage-specifying and signature cytokines, whereas CD27(-) γδ T cells showed bivalent marking at those key genes that associated with their differentiation into IL-17+ IFN-γ+ effector cells in vitro and in vivo [88].

Histone modification writers, readers, and erasers in lymphocyte differentiation

Studies on the writers, readers, and erasers of histone marks aimed to shed light on their functional impact in the establishment and maintenance of lymphocyte identities. Loss of regulators that set repressive marks to drive the plasticity of Th-cell subsets, pointing to a role of epigenetic mechanisms in inhibiting functional plasticity and therefore a crucial impact of histone modification in lineage commitment [82, 89]. The SUV39H1–H3K9me3–HP1a silencing pathway was shown to control the stability of Th2 cells [89]. Th2-cells deficient in SUV39H1A, the histone lysine methyltransferase setting repressive H3K9me, had a reduced ability to silence Th1-driving genes in Th2 cells and SUV39H1A-deficient Th2 cells expressed IFN-γ when re-cultured under Th1-driving conditions. Furthermore, conditional knockout of Ezh2 in CD4+ T cells, a key component of the polycomb repressive complex (PRC) 2 responsible for setting H3K27me3 marks, led to the spontaneous generation of IFN-γ and Th2 cytokines in non-polarizing conditions and enhanced functional plasticity in Th cells [82].

Ezh2-deficient Th2 cells were more prone to produce IFN-γ in re-polarizing conditions compared to WT cells. Mechanistically, Ezh2-mediated repression of master TFs like T-bet and Eomes was required to prevent IFN-γ production by Th cells [82]. Ezh2 was further shown to be important to regulate the stability of iTreg cells [90–93]. Loss of Ezh2 function in iTreg cells, characterized by Foxp3 expression, resulted in multi-organ autoimmunity and production of pro-inflammatory cytokines [92, 93].

Further support of a critical role for epigenetic regulators to maintain developmental programs comes from cross-differentiation experiments of different lymphoid fates. T cell-specific Ring1A/B-deficient mice have a block in T-cell development but those early arrested T cells can develop toward functional B cells when transferred to immunodeficient mice, with an additional deletion of Cdkn2a, a cell cycle inhibitor [94]. Ring1A and Ring1B are core components of PRC1 that mediate monoubiquitination of H2AK119. Histone H2AK119 monoubiquitynation leads to a conformational change of histones to a repressive state [94, 95]. Interestingly, the severe T-cell developmental block detected in Ring1a/b conditional deficient mice was almost overcome by additional deletion of Pax5, the key B-cell lineage specifying TF [94]. These results underline that Pax5 is a major target of polycomb proteins during early T-cell development and point to a crucial role of epigenetic suppression of the B lineage-specific gene program in T-cell fate establishment [94]. In sum, this study is the first to report a true lineage conversion in mice with genetically modified polycomb components, where previous studies targeting different polycomb genes detected a more general defect in the development of various types of hematopoietic cells [96–99].

Writers of active marks are implicated in maintaining lineage specifications [100]. Th2 cells with haploinsufficiency in MLL-4, a histone lysine methyltransferase setting H4K4me marks, show a defect in the maintenance but not the induction of *Gata3*, *Il4*, *Il5*, and *Il13* expression, and Th2 memory formation, whereas Th1-cell differentiation was not affected [100]. MLL-4 is thought to preferentially catalyze the monomethylation of H3K4 (H3K4me1) and occupies intergenic and enhancer regions [101]. A recent study highlights a role of MLL-4 in the induction but not maintenance of both thymic-derived Treg and in vitro-induced Treg cells, by affecting chromatin-configuration changes that are permissive for Foxp3 induction [102]. By using high-resolution chromatin-confirmation-capture protocols the study further proposed a role of MLL4 to promote looping of chromatin. H3K4me1 sites are established as a result of direct MLL4 binding, which were termed "anchor sites." These anchor points are looped into contact with distal regulatory chromatin segments and ensure multiple genomic locations where MLL4 can catalyze H3K4me1 without the need of being recruited to every single location [101, 102]. MLL4 indirect targets are recruited by MLL4, other components in the MLL4 complex or additional factors that interact with the MLL4 complex for the modification H3K4me1 [101, 102]. Subsequently, the established H3K4me1 sites prime the Foxp3 locus for transcriptional activation in Treg cells [102].

A downstream effect of histone modifications is their recognition by reader proteins. Research in the upcoming years has to elucidate the precise role of histone mark readers in immune cell differentiation. Some studies already point to a role during cellular differentiation. By employing a selective small molecule inhibitor (JQ1) against bromodomain and extraterminal (BET) proteins it was shown that BET proteins are important for the differentiation and activation of human and murine Th17 cells [103]. BET proteins recognize acetylated histones and consist of several members. Brd2 and Brd4 bind directly to *Il17* and regulate IL-17 itself as well as several additional Th17-related cytokines, like IL-21 and GM-CSF [103]. Using JQ1 in autoimmune mouse model like EAE and induced collagen-induced arthritis (CIA) showed its potency to reduce disease severity [103]. A knockout of mel18, a component of PRC1 that binds to H3K27me3, showed impaired Th2-cell differentiation, which associated with both decreased demethylation of the IL-4 gene and reduced expression of GATA-3 [104].

The coordinated removal of histone marks is another mechanism to regulate both developmental as well as functional differentiation. Histone deacetylase (HDAC) 7 is specifically expressed in lymphocytes and mediates the removal of active acetylation marks on histones. Conditional deletion of HDAC7 blocks early B-cell differentiation as HDAC7 expression is needed to repress lineage-specific genes of myeloid and T cells in early developing B-cell progenitors [105]. HDAC7 recruitment to enhancers of target genes is needed for removal of active histone marks and HDAC7 deficiency ultimately leads to increased cell death of pro- and pre-B cells [105]. Also in the T-cell lineage, early studies concentrated on inhibitors of histone deacetylases (HDACs) highlighted a role of histone deacetylases in restricting the expression of both IFN-γ and Th2-related cytokines [45, 106]. Additionally, HDAC6 together with sirtuin-1 was shown to control Treg function by deacetylating and reducing expression of Foxp3 [107].

In recent years, research focusing on the differentiation of germinal center B cells highlighted an important role of histone-modifying enzymes for coordinated lineage commitment. GC B cells have a striking change in their transcriptional output compared to naïve B cells and several classes of genes have to be repressed to establish their unique phenotype [108]. Those repressed genes are linked to terminal differentiation, DNA damage, and proliferation checkpoints. The histone lysine methyltransferase KMT2D, the acetyltransferases CREBBP, and P300 are involved in the coordinated changes that occur during GC formation and misregulations of these pathways are linked to lymphomagenesis [109–112]. Those histone-modifying enzymes are involved in either activation or repression of enhancers during GC B-cell differentiation through H3K27ac and H3K4me, respectively. A recent study highlights an additional role of the histone demethylase LSD1 in this process as a conditional deletion of *Lsd1* in GCs significantly impaired GC formation [113]. LSD1 specifically catalyzes the demethylation of H3K4me1/2 [114] and was previously implicated in early hematopoietic stem cells differentiation and terminal blood cell maturation [115]. Using a GC B-cell conditional knock out model of *Lsd1* elucidated a link of the TF Bcl6 to recruit Lsd1 to intergenic and intronic enhancers which ultimately affect the transcriptional output of GC B cells by changing chromatin accessibility [113]. LSD1 loss of function caused the failure to repress genes involved in GC exit and terminal differentiation which could explain short-lived GCs in Lsd1-deficient mice [113].

Collectively, these examples of histone-modifying enzymes or histone-modification readers underline an important role of these pathways in lymphocyte lineage commitment and maintenance.

Chromatin accessibility in the differentiation of immune cells

Similarly to DNA methylation and histone modifications patterns, chromatin accessibility can be used to identify distinct cell populations. The establishment of gene expression programs and ultimately cell identities depend not only on lineage-specific TFs and their cross talk within an epigenetic landscape but also on chromatin accessibility. Although not an epigenetic-mechanism per se, the packaging of DNA/nucleosomes within the nucleus leads to the organization of the genome into distinct spatial structures determining gene expression profiles by providing relative accessibility (open or closed) of key regulatory regions to TFs (Fig. 2). Several techniques are available to investigate chromatin accessibility including DNase-, ATAC-, Mnase-, and NOMe-sequencing (reviewed in Ref. [116]). The power of chromatin accessibility assays to define cell types regarding their epigenome is illustrated by studies using single-cell ATAC-seq in several organisms including mouse and drosophila tissues [117–119]. Recently, strong dataset was generated in immune cells that form a basis for our understanding of cis-regulatory networks in distinct immune cell populations [120–123]. The Immunological Genome Project profiled chromatin accessibility additional to the transcriptome of over 80 unique subtypes of murine immune cells populations [123]. Those subtypes include not only mature lineages but also progenitors along differentiation trajectories, including HSCs and MPPs and for the B-cell lineage pro-B cells, pre-B cells isolated from the bone marrow and several mature B-cell types (naïve, follicular, GC B cells) and for the T-cell lineage DN1-DN4 cells isolated from the thymus and CD4 +, CD8 + T effector subsets and subtypes of unconventional T cells like γδ and NKT subsets [123]. This massive resource is available for in-depth analysis of how cis-regulatory landscapes vary between closely related cell types and which cis-regulatory regions have implications in specific immune cell differentiation [123]. More cell type focused studies revealed the chromatin changes during B- and T-cell differentiation and decipher crucial roles for lineage-determining transcription factors in instructing the establishment of a cell lineage, like TCF1 for T cells and EBF1 for B cells [121, 122]. Additionally, another study concentrated on the chromatin changes during CD8 + T-cell responses to acute and chronic viral infection [120]. Changes in accessible regions were identified in differentiation states of naïve, effector, and memory cells and distinct TF binding patterns were monitored. The comparison of different CD8 + T-cell subsets and NK cells during immune responses revealed shared accessible chromatin regions between naïve NK cells and memory CD8 + T cells, pointing to similar characteristics of these cells like rapid activation potential and cytolytic functions [124]. Through the combined analysis of shared accessible regions and common transcriptional changes in NK and CD8 + T-cell memory, a signature of genes has been identified that may play a role in the generation of both innate and adaptive lymphocyte memory [124]. Collectively, studies on the dynamics of chromatin structure in different cell states provide useful information on the capacity and plasticity of immune cells to respond to developmental or environmental cues by enabling TFs binding to specific location and establishment of unique transcriptomes.

Part II: RNA-based mechanisms of regulation of lymphocyte differentiation

The classical chromatin-based mechanisms of epigenetic regulation previously described have, in the last decade, been followed up by the increasingly explored world of noncoding RNAs. Two of the most functionally relevant groups of noncoding RNAs are long noncoding RNAs and microRNAs. These provide a new layer of genome expression regulation by acting mainly at the posttranscriptional level of gene regulation in diverse biological systems including lymphocytes [125–129].

MicroRNA-mediated regulation of T- and B-cell differentiation

MicroRNAs (miRNAs) are single-stranded RNAs conserved throughout the phylogenetic tree and their exact number is still not fully estimated since the identification of new miRNAs is continuously occurring. The latest release (v22 March 2018) of the miRBase database (http://www.mirbase.org), a primary miRNA sequence repository, includes 271 organisms. In the *Mus musculus* species, 1978 mature miRNAs (originated from 1234 precursor miRNAs) have been identified and, in *Homo sapiens*, 2654 mature miRNAs (originated from 1917 precursors) have been described.

miRNAs are generated from endogenous transcripts that are transcribed by RNA polymerase II in long dsRNA precursor transcripts designated by primary-miRNA (pri-miRNA), that are capped and polyadenylated [130, 131]. The primary miRNA precursors are processed by the ribonuclease (RNase) III Drosha-DGCR8/Pasha nuclear complex that originates smaller precursor miRNAs of 60–100 nucleotides that possess a hairpin structure. These pre-miRNAs are transported to the cytoplasm, through nuclear pores by the system of exportin5-RanGTP [132]. There, they are further cleaved by the Dicer enzyme into ~22-nt double-stranded mature miRNAs, (miRNA–miRNA* duplexes, where miRNA was defined as the antisense or guide/mature strand; and miRNA* the sense, or passenger strand). Helicases separate both miRNAs strands and the mature strand is loaded into the RNA-induced silencing complex (RISC) a hetero-oligomeric complex integrating an Argonaute (AGO) protein, Dicer and a dsRNA-binding protein (the TRBP in humans). Only one strand associates with the AGO protein becoming the guide strand, the non-guide strand is cleaved during loading of mature miRNA strand into RISC [131, 133]. The strand with the less stably paired 5′ end is the preferentially chosen to be loaded into the AGO proteins [134]. Finally, the guide strand drives the RISC complex toward the target RNA that has a nucleotide sequence complementary to that of miRNA (so-called "seed" sequence or short sequence at nucleotides 2–8 on the 5′end of the miRNA typically bind to the 3′ UTRs of their target mRNAs), and after pairing the mRNA is degraded or its translation abolished [135, 136]. Of note, it is estimated that approximately 85% of miRNA-mediated regulation in mammals occurs at the level of mRNA decay [137].

While the global importance of miRNAs is clearly illustrated by the developmental failure of Dicer-deficient embryonic stem cells (in vitro) and embryos (in vivo) [138], specific depletion of Dicer in early stages of immune cell differentiation, as well as unique spatial and temporal expression patterns in distinct lymphoid (and myeloid) lineages, are clearly suggestive of multiple roles for miRNA in lymphoid development/differentiation and immune responses [125–128].

miRNA-deficient T cells generated by specific genetic inactivation of either Drosha, DGCR8 or Dicer exhibit reduced proliferation and survival after in vitro stimulation, as well as an increase in interferon-γ (IFN-γ) production and in the frequency of IL-17A-producing CD4+T cells, implying that miRNAs regulate both Th1 and Th17 cell differentiation [139–141]. Another effector T-cell population affected by miRNA depletion is the CD8+ T-cell population, whose numbers are markedly diminished in the periphery upon thymic deletion of Dicer and whose effector response is impaired upon deletion of Dicer in mature CD8+ T cells, namely, there is reduced survival of antigen-specific effector CTLs during viral or bacterial infections [141, 142]. In contrast with these in vivo observations, deletion or depletion of Dicer in mouse or human activated CD8+ T cells in vitro causes upregulation of perforin, granzymes, and effector cytokines [143]. Indeed, specific miRNAs were shown to impact CD8+ T-cell differentiation and can explain both the in vivo and the in vitro observations as detailed below.

Specific deletion of Dgcr8D in the presence of IFN-γ blocking antibodies and Th2-cell polarizing conditions, increased the frequency of IL-4+ and IL-13+ cells, indicating that miRNAs also play a role in Th2-cell differentiation [144]. Tfh cell differentiation is also compromised in the absence of miRNAs. Mice deficient in the miRNA biogenesis factor DGCR8 exhibit a diminished generation of Tfh cells resulting in lower relative and absolute numbers of GC B cells. In addition, due to the loss of capacity to upregulate CXCR5 expression, miRNA-deficient Tfh cells fail to accumulate at the vicinity of B cells at the interface between the T- and B-cell zones, not entering into B-cell follicles. [145]. On the regulatory T-cell counterparts, it was observed that Dicer-deficient Treg cells fail to keep their signature gene expression program, downregulated Foxp3 and upregulated IL-4 and IFN-γ expression and lost suppression activity in vivo, thus leading to a fatal autoimmune phenotype [146, 147]. Finally, the iNKT developmental program is clearly dependent on miRNAs since the specific deletion of Dicer in thymocytes leads to a selective impairment of iNKT cell survival and functional differentiation [147, 148].

B-cell differentiation has also proven to be dependent on miRNAs, as observed in several studies in which Dicer was specifically deleted at the earliest stage of B-cell differentiation (Dicerfl/fl mb1-Creki/+ mice) [149]. These Dicer-deficient mice displayed a severe block of B-cell development at the pro-B to pre-B transition, at least partly due to massive apoptosis of Dicer-deficient pre-B cells, which resulted in an almost complete absence of B cells in the periphery [149]. Moreover, Aicda-Cre specific deletion of Dicer in activated B cells impaired the formation of GC B cells [150]. Upon Dicer deletion at later stages, using a CD19-Cre Ki/+ Dicerfl/fl mouse model [151], it was possible to observe a decrease in the generation of mature splenic and lymph node B cells, whereas bone marrow differentiation was unaltered.

Overall, these studies have shown that the differentiation of B and T cells is regulated by miRNAs, thus prompting the dissection of specific miRNAs involved in these processes.

Specific miRNAs regulating T- and B-cell differentiation

The unequivocal importance of miRNAs for T-cell differentiation has been corroborated by the growing evidence that miRNAs are an integral part of gene expression networks, determining T-cell identity, and function by posttranscriptional repression of target mRNAs, including those involved in T helper cell polarization, CD8+ T-cell functions and NKT-cell differentiation [145, 146, 152, 153]. Thus, miRNAs seem to act mainly as regulators of key cytokines or transcription regulators, either promoting or inhibiting T-cell differentiation/function. Although explored to a lesser extent, specific miRNAs are also clearly required to B-cell differentiation, impacting on different stages of B-cell development.

Table 1 summarizes some of the most relevant miRNAs thus far implicated in T-cell differentiation (partially reviewed in [208]); their main targets, altered expression levels and physiological relevance are indicated. To name a few examples, miR-29 acts by directly targeting T-bet and Eomesodermin (Eomes), two TFs that regulate IFN-γ expression [139], and also IFN-γ mRNA itself [154], thus blocking Th1 differentiation; miR-24 targets the 3′UTR of IL-4 to suppress its expression and inhibits Th2-cell differentiation [144] and the miR-106∼363 cluster that is downregulated upon Th17 differentiation in mice, and its members miR-18b, miR-106a, and miR-363-3p target the Th17-associated TF ROR-α, as well as nuclear factor of activated T cells (NFAT), thus decreasing Th17 differentiation and IL-17A production [168] miR-146a is considered as a “brake” in T-cell activation and differentiation as it acts by dampening the nuclear factor kb (NF-kB) activation [209, 210]. In its absence, an enhanced NF-kB activation was observed that was linked to higher proliferation and increased production of effector cytokines including IFN-γ, IL-17A, and IL-2 by CD4+ and CD8+ T cells, thus leading to a general activation phenotype [209]. Conversely, those miRNAs promoting T-cell differentiation usually targets inhibitors of a given pathway relevant for this process. An example is the miR-17∼92 (mainly miR-19b) cluster that targets PTEN (phosphatase and tensin homolog), a negative regulator of TCR signaling, more specifically of the PI3K-AKT-signaling pathway, affecting Th1-cell differentiation and IFN-γ production [160], Th2-cell differentiation [167], Tfh cell differentiation [189], and CD8 T-cell expansion [200]. PTEN is also targeted by miR-181, which was reported as an essential regulator that supports the proliferative developmental stages of NKT cells [204, 205]. Overall these experiments indicate that the same miRNAs can act in different T-cell populations either via the same target mRNA or via distinct targets.

miRNAs can also act in a more indirect manner: for example, miR-125a promotes Treg activity by suppressing effector T-cell factors, such as STAT3, IFN-γ, and IL-13, which stabilizes the immunosuppressive capacity of Treg cells [183]. On the other hand, miR-10a also limits lineage conversion of Treg cells by preventing the acquisition of Th17 and follicular helper T-cell features [211]. Focusing on γδ we have shown that

Table 1 miRNAs that regulate T-cell differentiation.

T cell	miRNA	Target mRNA and general mechanism	Altered expression levels and physiological relevance	Reference
Th1	***miR-29***	T-bet, Eomes, IFN-γ Critical targets for Th1-cell expression program	Decreased miR-29 expression in MS patients contributes to chronic inflammation Mice with transgenic expression of a 'sponge' to miR-29 (GS29) has increased Th1 responses and higher resistance to infection with BCG or *Mycobacterium tuberculosis*.	[139, 154, 155]
	miR-140-5p	STAT1 Promotes T-bet expression	miR-140-5p Expression is inversely correlated with development of MS	[156]
	miR-146a	PRKCε Phosphorylates STAT4 promoting Th1 cell differentiation	Reduced miR-146a expression levels and increased PRKCε levels is observed in early hyperinflammatory phase of sepsis	[157]
	miR-10a	NOD2, IL-12/IL-23p40 Dendritic cell IL-12/IL-23p40 and NOD2 expression promotes IL-12 and IL-23 production inducing Th1 (and Th17) cell differentiation	miR-10a expression is decreased in the inflamed mucosa of IBD, downregulating the mucosal inflammatory response.	[158]
	miR-155	SHIP1 Negative regulator of PI3k pathway	Promotes immune responses, acting as a regulator of CD4 + (and CD8 +)T cell-mediated antitumor immunity.	[159]
	miR-17∼92 cluster	PTEN Negative regulator of PI3k-mTOR signaling pathway	Promotes anti-viral IFN-γ responses Loss of miR-17-92 in CD4 T cells results in tumor evasion	[160]
	miR-24	TCF1 Inhibits IFNγ expression	Elevated miR-24 expression levels were detected in patients with ulcerative colitis and rheumatoid arthritis being potential biomarkers of these diseases	[161–163]

Continued

Table 1 miRNAs that regulate T-cell differentiation—cont'd

T cell	miRNA	Target mRNA and general mechanism	Altered expression levels and physiological relevance	Reference
Th2	***miR-23∼24∼27 cluster***	TCF1 Induces GATA3 expression IL-4 Promotes Th2 cell differentiation GPR174, cyclin K and AF4/FMR2 GATA3, IKZF1 and NFATC2	Mice lacking both miR-24 and miR-27 clusters in T cells have increased Th2-cell responses and tissue pathology in a mouse model of asthma.	[144, 161, 164]
	miR-155	S1pr1 Binds to S1P and regulates lymphocyte maturation, migration and trafficking.	S1pr1 mRNA and protein expression was increased in PBMCs from SLE patients miR-155$^{-/-}$ mice exhibit reduced airway inflammation with reduced eosinophilia compared to wt counterparts.	[165, 166]
	miR-19a (from miR-17-92 cluster)	PTEN, SOCS2 A20	miR-19a is Up-regulated in T cells of airways of patients with asthma. Cells lacking the miR-17∼92 cluster have markedly impaired Th2 responses.	[167]
Th17	***miR-24 (from miR-23-23-27 cluster)***	TCF1 – transcription factor that restricts Th17 (and Th1) responses	miR-24 expression detected in patients with ulcerative colitis and rheumatoid arthritis	[161–163]
	miR-10a	NOD2, IL-12/IL-23p40 Dendritic cell IL-12/IL-23p40 and NOD2 expression promotes IL-12 and IL-23 production inducing Th17 (and Th1) cell differentiation	miR-10a expression is decreased in the inflamed mucosa of IBD, downregulating the mucosal inflammatory response.	[158]
	miR-106a-363 cluster	ROR-α, NFAT Th17 transcription factors	Not determined	[168]
	miR-18a	Smad4, Hif1a and Rora Key transcription factors in Th17 expression program	miR-18a–deficient mice displayed increased Th17 responses in airway inflammation models in vivo.	[169]

miR-146a	TRAF6, IRAK1 Adaptor molecules involved in the NF-κB activation pathway	Up-regulated in CD4 + T cells of EAE mice Blocks autocrine IL-6/IL-21 Th17 differentiation pathway Mice with deficient miR-146a T cells have more severe EAE.	[170]
miR-326	Ets-1 Negative regulator of Th17 cell differentiation	miR-326 expression levels are correlated with disease severity in EAE mice and MS patients	[171]
miR-223	Unknown	Upregulated in EAE pathogenesis	[172]
miR-384	SOCS3 Inhibitor of IL-6-JAK-STAT3 pathway)	Upregulated in EAE pathogenesis	[173]
miR-590	Tob1 Member of the tob/btg1 family of anti-proliferative proteins.	miR-590 is upregulated in PBMCs and cerebrospinal fluid of MS patients Significantly decrease in the expression of Tob1 is observed in sera and CSF of patients with MS correlating with demyelinating disease activity.	[174, 175]
miR-448	PTPN2 Anti-inflammatory player with capacity to suppress Th17 differentiation.	miR-448 levels are increased expression in MS patients and in Th17 cells PTPN2 expression levels are decreased in PBMC and CSF of MS patients	[176]
miR-155	Jarid2 Histone demethylase, which promotes the recruitment of the Polycomb Repressive Complex 2 (PRC2) to chromatin, and mediates transcriptional regulation of cytokine genes in Th17 cells.	miR-155 levels are increased in MS patients, with upregulation of Th17 cytokines, compared to healthy donors. MiR-155 is overexpressed in DSS-induced mice, while Jarid2 is downregulated.	[177, 178]

Continued

Table 1 miRNAs that regulate T-cell differentiation—cont'd

T cell	miRNA	Target mRNA and general mechanism	Altered expression levels and physiological relevance	Reference
	miR-873	Foxo1 TF inhibitor of RORγT	miR-873 expression levels are increased in PBMCs from SLE patients. lentivirus-mediated inhibition of miR-873 diminishes disease severity of spontaneous SLE in mice.	[179]
	miR-425	Foxo1 TF inhibitor of RORγT	miR-425 is upregulated in PBMCs and mucosal cells of patients with IBD. Inhibition of miR-425 significantly reduced the disease severity of TNBS-induced colitis in mice	[180]
Treg	***miR-31***	Gprc5a Retinoic acid-inducible protein 3	Gprc5a$^{-/-}$ mice developed EAE earlier and more severely than Gprc5a$^{+/+}$ mice. Deficiency in Gprc5a causes impaired pT$_{reg}$ cell induction	[181]
	miR-27	c-rel, GZMB, FOXO1, RUNX1, SMAD2/3 and IL-10	Tregs with overexpressing miR-27 exhibited diminished homeostasis and suppressor function in vivo	[182]
	miR-125	STAT3, IFN-γ and IL-13	Mice deficient for miR-125a show a shift from immune suppression to inflammation, presenting more severe pathogenesis of colitis and EAE	[183]
	miR-155	SOCS1 Negative regulator of IL-2	Not determined	[184]
Tfh	***miR-146a***	ICOS Provides costimulatory signals for optimal Tfh and GC B-cell differentiation	miR-146a deficiency in T cells initiates Tfh and GC B-cell accumulation. This phenotype is enhanced by T-cell extrinsic factors.	[185]
	miR-155	Peli1 Ubiquitin ligase that promotes degradation of c-Rel and Fosl2, among 20 other genes.	Analysis of lymphocytes from inflamed miR-146a−/− mice have shown a miR-155-dependent phenotype of elevated numbers of Tfh cells, GC B cells, and autoantibodies.	[186, 187]

	miR-17-92 cluster	PTEN Negative regulator of PI3k-mTOR signaling pathway PHLPP2 and RORA	T-cell specific deletion of miR-17-92, impaired Tfh-cell differentiation and GC formation and fails to control LCMV infection. Overexpression of miR-17-92 induces large numbers of Tfh cells and lymphoproliferative disease and autoimmune SLE-like symptoms.	[145, 188, 189]
CD8 T cells	***Let-7***	Myc, Eomes Transcription factors that activate CD8 differentiation program	Let-7 expression impairs the clonal expansion and differentiation of CTLs in response to viral infection in vivo.	[190]
	miR-146a	STAT1 Promotes T-bet expression	miR-146a levels are increased in both CD4$^+$ and CD8$^+$ T cells of patients infected with HBV	[191]
	miR-491	CDK4, TCF-1 and Bcl2l1/Bcl-xL	miR-491 levels are increased in spleen CD8$^+$ T cells from tumour-bearing mice.	[192]
	miR-23a	Blimp-1 Transcription factor that promotes CTL toxicity and effector cell function	miR-23a is upregulated in tumour-infiltrating CTLs of advanced lung cancer patients.	[193]
	miR-155	SHIP1, Akt, SOCs1 and Ptpn2	Required for CD8 + T-cell responses to both viruses and cancer cells.	[159, 194–197]
	miR-150	c-Myb Tanscription factor that controls lymphocyte differentiation and proliferation.	miR-150 deficient CD8 + T cells provide better protection during secondary pathogen challenge. Myb overexpression increases CD8$^+$ T-cell memory formation, polyfunctionality and recall responses that promoted curative antitumor immunity after adoptive transfer.	[198, 199]
	miR-17-92 cluster	PTEN Negative regulator of PI3k-mTOR signaling pathway	knock out of miR-17-92 with GzB-cre reduces the number of LCMV-specific CD8$^+$ T cells.	[200]

Continued

Table 1 miRNAs that regulate T-cell differentiation—cont'd

T cell	miRNA	Target mRNA and general mechanism	Altered expression levels and physiological relevance	Reference
NKT	***Let7***	PLZF NKT-cell lineage specific transcription factor	The upregulation of let7 miRNAs in NKT cell development downregulates PLZF and promotes their terminal differentiation to IFN-g producing NK1 cells. The downregulation of let7 maintains high PLZF expression and terminally differentiate NKT cells into IL-4-producing NKT2 and IL-17-producing NKT17 cells.	[201]
	miR-155	ETS1, IKT	miR-155 overexpression caused a block of iNKT cell maturation in the thymus reflected in an overall reduction of peripheral iNKT cells.	[202]
	miR-150	c-Myb Transcription factor controlling lymphocyte differentiation and proliferation.	c-Myb expression is upregulated in immature and mature iNKT cells from miR-150KO mice when compared to WT mice.	[203]
	miR-181	PTEN Negative regulator of PI3k-mTOR signaling pathway	miR-181-deficient mice possess defects in lymphoid development and T-cell homeostasis associated with impaired PI3K signaling	[204, 205]
	miR-133b	Th-POK Zinc finger transcription factor that acts a key negative regulator of thymic NKT17 cell differentiation.	By regulating NKT17 cell function, miR-133b might play a role in infections and in autoimmune diseases	[206]
γδ T cells	***miR-146a***	Nod1 Intracellular pattern recognition receptor that enhances IFN-γ production in αβ T cells	Nod1-deficient mice lack IL-17 + IFN-γ + γδ27- cells and do not resist to *Listeria monocytogenes* infection.	[207]

Bold italic miRNAs that repress differentiation.
Bold miRNAS that promote differentiation.

miR-146a, which is highly expressed in the CD27(-) γδ T-cell subset that produces IL-17, limits IFN-γ production by this subset, thus restricting its functional plasticity both in vitro and in vivo [207].

Although explored to a lesser extent than in T-cell differentiation, miRNAs are also unequivocally required for B-cell differentiation. As previously mentioned, mice deficient in Ago2 and Dicer fail to develop pro-B cells into pre-B cells, which is critical for B-cell lineage commitment [212]. Several miRNAs have been reported to specifically affect different stages of B-cell development. While miR-181a and the miR-17-92 cluster positively regulate this process, miR-34a, miR-150, and miR-212/132 negatively regulate B-cell development [213–220].

miR-181a, a key miRNA in the modulation of the thresholds of TCR signaling in thymocytes [221, 222], was in fact firstly reported to affect B-cell differentiation in the bone marrow [213]. Overexpression of miR-181a by retroviral transduction in HSCs significantly increased the differentiation toward B-cell lineage both in vitro and in vivo [213], while miR-181ab1 deficient mice showed a mild decrease in the number of peripheral and germinal center (GC) B cells [223]. miR-181a is thus a good example of a miRNA playing distinct roles in different cell lineages (further explored below for other miRNAs), very likely by affecting various sets of target genes. Like miR-181a, the miR-17-92 cluster, [224], also promotes pro-B to pre-B stage differentiation. miR-17-92-deficient mice die shortly after birth from ventricular septal defects and lung hypoplasia [214] as a consequence of inhibition of B-cell development and, although controversial, likely via misregulation of the proapoptotic protein Bim [225].

miR-34a, miR-150, and miR-23a cluster, in, turn, negatively regulate B-cell development. mir-34 downregulates FOXP1, a B-cell oncogene [215], while miR-150 acts through its primary target c-Myb, a transcription factor controlling multiple steps of B-cell development. Upon miR-150, overexpression, c-Myb is inactivated leading to a partial block at the pro-B to pre-B transition in the bone marrow and to a decrease in conventional follicular cells and a subset of B cells that are responsible for the production of natural antibodies [151, 213, 226]. During late B-cell maturation in follicles, the downregulation of miR-150 is required for GC selection and development of the adaptive humoral immune response [219]. Finally, the miR-212/132 cluster suppresses B-cell differentiation at the pre-B to pro-B transition stage by repressing SOX4 [220].

Some miRNAs act on both T and B cells. In fact, as Tfh cells are critical determinants of antigen-specific B-cell immunity [227], miRNAs regulating this T-cell population can also impact on B cells. In this sense, miR-146a and miR-155 have been recently described as important posttranscriptional regulators with opposing functions on Tfh cell differentiation and B-cell differentiation. miR-146a directly targets ICOS in all T-cell subsets and maturation stages, and ICOS overexpression caused by miR-146a deficiency lead to Tfh cell accumulation [185]. In contrast to the immunosuppressive function of miR-146a, miR-155 acts as an important promoter of inflammatory responses in various cell types, including B cells, T cells, DCs, and macrophages. Thus, the ratio between miR-155 and miR-146a levels may be important for the regulation of immune responses.

The emerging role of lncRNAs in immune cell differentiation

Long noncoding RNAs (lncRNA) are regarded as unconventional epigenetic factors that can provide locus-specific control of gene expression. lncRNA comprise a large and heterogeneous class of transcripts, arbitrarily defined as being more than 200 nucleotides in length that are transcribed and processed similarly to mRNAs. The majority of lncRNAs is transcribed by RNA Polymerase II and is similarly subjected to splicing, polyadenylation, and 5′ capping. However, they generally exhibit lower expression, a fewer number of exons, and a much higher tissue specificity than mRNAs [228, 229]. The lncRNAs can be transcribed from intergenic regions, promoter regions or be interleaved, overlapping or antisense to annotated protein-coding genes [230]. Although the first report of a genome-wide lncRNA transcript expression profile occurred in 2011 [231], only a small portion of that has been molecularly studied. The most recent estimate by the Encyclopedia of DNA Elements (ENCODE) Project Consortium (GENCODE release 29) is that the human genome contains ~16,066 lncRNA genes that encode more than 29,000 distinct lncRNA transcripts. In what refers to the mouse genome (GENCODE release 20), it contains ~13,002 lncRNA genes that encode approximately 18,000 lncRNA transcripts. The majority of the lncRNAs (70%) are poorly conserved across species; and only a few lncRNAs show high sequence conservation across species, for example, XIST, PVT1, MIAT, NEAT1, MALAT1, and OIP5-AS [232]. Still, globally, lncRNA genes show higher sequence conservation than randomly selected genomic regions, including positional as well as promoter regions conservation, splice sites or the act of transcription itself (reviewed in [232]).

The biological significance of lncRNAs was proven in a murine knockout study in which developmental defects and lethality occurred upon deletion of several lncRNAs [233]. A recent large-scale lncRNA knockout screening in human cell lines identified a significant effect on cancer cell growth for 50 out of 700 lncRNAs tested [234]. Thus, although the functional significance of the bulk of lncRNAs is unclear, a significant proportion of them have biological functions.

The functional versatility of lncRNAs includes chromatin modification, nuclear domain organization, transcriptional control, regulation of RNA splicing and translation, and modulation of protein activity (reviewed in Ref. [235]).

LncRNAs can, in general, regulate the expression of single or multiple genes. In the former case, genes encoding lncRNAs are often in close proximity in the genome to the protein-coding gene the lncRNA regulates and are coexpressed in a "guilt-by-association manner." Although this can also apply to the latter case, expression of multiple protein-coding genes is often regulated by the lncRNA via direct or indirect mechanisms, with these genes likely being scattered throughout the genome. Thus, an examination of lncRNA and mRNA coexpression across multiple biologic samples can provide insight into the potential targets of a given lncRNA. Moreover, lncRNAs often interact with proteins of the chromatin structure, thus providing an extra layer of regulation not observed with other types of noncoding RNAs.

The study of lncRNAs in the immune system/response is an emerging field. Relatively few lncRNAs have been associated with the immune response and even fewer have been thoroughly explored functionally and mechanistically. However, the study in this setting is very promising as the immune system provides a highly organized biological context in which cellular phenotypes and functions are finely mapped, cellular components are easily accessible and manipulated, and perturbation at the molecular and cellular levels can be achieved through in vitro and in vivo models.

Although several lncRNAs were shown to play a role in a variety of immune responses, including response to infections by macrophages and other innate cells, we will focus here on the role of lncRNA in T and B-cell differentiation. While the role of lncRNAs has been less explored in B-cell development, when compared to T-cell differentiation, considerable insights have been provided based on expression profile and correlation studies. In contrast to mRNA expression patterns, lncRNA expression patterns can distinguish cells committed to the B- and T-cell lineage already at the progenitor stage in bone marrow [129], indicating their importance in lymphoid lineage commitment.

LncRNAs in T-cell differentiation

Genome-wide analysis of lncRNA expression indicates that hundreds of lncRNAs are induced by inflammatory stimuli [236, 237] and that thousands are induced across various stages of T-cell development and activation [238]. T-cell stage and lineage-specific expression of lncRNAs and mRNAs have been studied in both humans and mice. It was observed, in general, that, unlike mRNAs, most lncRNAs are expressed in a stage or lineage-specific manner. This can be seen across T-cell development in the thymus and T-cell differentiation in the periphery. Furthermore, these T-cell–specific lncRNAs are not expressed, or are present at very low levels, in nonlymphoid tissues. Thus, these cell types can be defined by unique lncRNA expression profiles [238–240].

Various lncRNAs have been implicated in Th1, Th2 and Th17 cell differentiation and their functional properties. LncRNA Tmevpg1, also named NeST, lncR-Ifng-3′ AS or IFNG-AS1, is expressed by the Th1 subset in a T-bet-dependent manner and is necessary for the efficient transcription of *Ifng* in Th1 cells [241]. In CD8+ T cells, Tmevpg1 associates with the WDR subunit of mixed-lineage leukemia histone H3 Lys 4 (H3K4) methyltransferases to increase H3K4 methylation at the *Ifng* locus, being required to confer resistance to lethal *Salmonella enterica* infection in mice, thus establishing an essential role for this lncRNA in the host response to a bacterial pathogen [242].

In mouse and human Th2 cells, it was observed that GATA-3, a key transcription factor of the Th2 program, is coexpressed with the lncRNA GATA3-AS1 [238]. GATA3-AS1 is a divergent lncRNA, transcribed in the opposite direction from GATA-3 in both species. As such, it is postulated that GATA3-AS might play a role in allergic or asthmatic responses in humans that typically induce a Th2-cell response. Th2 cells also express the lncRNA lincR-Ccr2-5′AS, which is regulated by GATA-3. Knockdown of lincR-Ccr2-5′AS has revealed that it controls the expression of nearly 1200 genes which overlap with genes dependent on GATA-3 [238]. In particular, lincR-Ccr2-5′AS upregulates a cluster of genes encoding key Th2-cell chemokine receptors (*Ccr1*, *Ccr2*, *Ccr3*, and *Ccr5*). Moreover, compared with controls, Th2 cells from which lincR-Ccr2-5′AS is depleted display an impaired ability to migrate to the lungs after in vivo transfer [238]. Another independent study of human T cells has also identified a cluster of antisense lncRNAs transcribed from the RAD50 locus (which encodes a double-strand-break repair protein) that are coexpressed with and regulate the transcription of the neighboring gene cluster encoding IL4, IL5, and IL13 genes under Th2-polarizing conditions [239].

The differentiation of Th17 cells was also shown to be dependent on the activity of a lncRNA, named Rmrp ('RNA component of the mitochondrial-RNA-processing endoRNase'). Rmrp was shown to associate with the DEAD-box RNA helicase DDX5, a functional partner of the major IL-17A transcription factor RORγ [243]. Th17 cells generated in vitro from DDX5-deficient mice produce less IL-17A than wild-type cells, consistently with the failure to in vivo- induced Th17 cell-driven organ inflammation upon transfer of DDX5-deficient T cells in a mouse model of autoimmune colitis. The function of DDX5 was shown to be dependent on interaction with lncRNA Rmrp through its RNA-helicase domain. In Th17 cells, Rmrp localizes to the nucleus and promotes assembly of the RORγt–DDX5 complex at genomic loci of genes encoding critical Th17 cell effector molecules, including *Il17a* and *Il17f*. Interestingly, knockdown of Rmrp RNA in human T cells leads to the compromised secretion of cytokines from Th17 cells, and mutant forms of Rmrp cause cartilage-hair hypoplasia, a congenital disease associated with immunological dysfunction [244, 245].

Like miRNAs, also lncRNAs seem to play a role in the balance between T effector and T regulatory cells. When MALAT1 (a lncRNA decreased in the spinal cords of EAE mice) expression was downregulated by RNA interference, there was a shift in the pattern of T-cell differentiation toward a Th1/Th17 profile and decreased Treg cells. The proliferation of T cells was also increased following MALAT1 downregulation, thus pinpointing a potential anti-inflammatory effect for MALAT1 in the context of autoimmune neuroinflammation [246].

Interestingly, some lncRNAs from one program can repress transcription factors from other programs, thus promoting a particular differentiation fate. Such is the case of MAF-4, which is expressed in Th1 cells, represses the expression of the Th2-cell transcription factor MAF to promote T-cell differentiation toward the Th1-cell lineage [240].

Together, these studies on T helper cells have demonstrated that lncRNAs can serve as critical regulators of cell-type-specific effector programs, often in concert with lineage-specifying transcription factors.

LncRNAs in B-cell differentiation

A large number of lncRNAs are expressed in a cell-type/stage-specific manner across B-cell differentiation, consistent with studies in T cells and other tissues, and suggesting a cis-regulatory function for many of them [247–250]. Compared to protein-coding genes, lncRNAs exhibit significantly less sequence conservation, but alternative comparative methods, such as structure, profile, or positional conservation can detect homologues without primary sequence conservation [232].

lncRNA RP11-132N15.3, located downstream of *Bcl6* [251] is an example of a lncRNA with a murine ortholog showing both sequence and positional conservation. However, this lncRNA appears to be downregulated in murine GC B cells [248], while being upregulated in human GC B cells (consistently with high BCL6 expression in GC B cells) [251], indicating that despite strong similarities a functional conservation is unlikely. In the same study, it was possible to identify lncRNAs specifically expressed during different stages of human B-cell development [251]. Notably, among sequences expressed mainly in pre-B cells, several of the identified lncRNAs included antisense transcripts (*SMAD1-AS1*, *MYB-AS1*, and *LEF1-AS1*) to transcription factors known to have a function in early B cells. Although the role of these lncRNAs in B-cell development is still unknown, their crucial positions in the gene coexpression network analysis suggest key functions in the early stages of cellular differentiation.

CRNDE (colorectal neoplasia differentially expressed) was another lncRNA specifically enriched in pre-B and also on GC CBs (centroblasts), coinciding with the overexpression of genes involved in processes related to mitosis and cell cycle control and consistently with the high proliferative activity of these cells [252]. Based on a study showing that in tumor cells *CRNDE* expression favors the metabolic switch to aerobic glycolysis [253], which is required during rapid cell proliferation, its expression principally in immature B cells might be consistent with its function as a metabolic regulator.

A study in mice reported that expression of paired box 5 (PAX5), a transcription factor that is crucial for B-cell commitment [254], led to differential expression of several lncRNAs, including enhancer-associated lncRNAs, which were shown to be bound by PAX5, and for which human orthologs have been described [248].

Another study in mice proposed a dominant role of germ-line transcribed lncRNAs during V(D)J recombination in progenitor B cells [255]. The most abundant transcripts were the PAX5-activated intergenic repeat (PAIR) elements PAIR4 and PAIR6, which are transcribed antisense to PAX5. These lncRNAs are essential for locus compaction and optimal positioning of neighboring heavy chain genes optimally for gene rearrangements to occur [255]. Remarkably, B cells that are deficient of the transcription factor YY1, which is necessary for distal VH gene rearrangements and precursor B-cell transition [256], displayed a marked reduction in both antisense transcription and DNA looping between the PAIR promoter and the intronic enhancer, compared to B cells with intact YY1 [255] supporting a pivotal role of PAIR4 and PAIR6 in V(D)J recombination during B-cell development. YY1 has also been proposed to interact with and relocate the lncRNA Xist, to the inactivated X-chromosome in activated B cells, thereby changing the X-linked gene regulation in these cells compared to antigen naïve B cells [257].

At later stages of B-cell development and maturation, lncRNA expression profiles can be highly similar between functionally distinct B cells such as follicular and marginal zone B cells in spleen [248] and naïve and memory cells in tonsils [250]. Only the strongly proliferative GC B cells showed a lncRNA profile that is very distinctive from other mature B-cell subsets [129, 248] suggesting than in B-cell lncRNAs may be most relevant in the early stages of B-cell differentiation.

Cross talk between noncoding RNAs

The different layers of epigenetic regulation, from DNA and histone modifications to chromatin remodeling and noncoding RNA-mediated regulation communicate with each other. LncRNAs for example, in addition to directly regulating transcription, also modulate chromatin through the specific recruitment of histone and chromatin-modifying complexes on one hand and by recruitment of transcription factors on the other hand. Specifically, within noncoding RNAs, the interplay occurs mostly as a consequence of sequence complementarity. On the one hand, miRNAs could bind to lncRNAs and promote their decay, similarly to mRNAs, on the other hand, lncRNAs can sequester miRNAs away from mRNAs, functioning as "miRNA sponges" or "miRNA decoys" and preventing them from acting on the protein-coding mRNAs. In addition, some lncRNAs might themselves encode miRNAs, both noncoding RNAs acting in a similar way (Fig. 2).

Some noncoding RNAs were shown to have a role in chromatin remodeling both in T- and B-cell differentiation. An example of the former is linc-MAF-4, which, as previously mentioned, represses the expression of the transcription factor MAF in Th2 cells to promote Th1-cell lineage commitment [240]. This is a consequence of long-distance chromosome contacts between genomic regions of linc-MAF and MAF, which allow linc-MAF-4 to recruit the chromatin remodelers EZH2 and LSD1 that in turn place repressive chromatin marks on the *MAF* promoter and repress its transcription. Consistent with these observations, linc-MAF-4 expression is significantly increased in peripheral blood mononuclear cells of patients with MS compared to control patients, its expression promoting activation of CD4+ T cells from patients with MS and correlating with relapse in patients with MS [240]. In B cells, the loss of the repressive lncRNA Xist from the inactive X(Xi) chromosome at the pro-B-cell stage, leads to loss of heterochromatin and reexpression of immune genes from the Xi. Activation of mature B cells restores Xist RNA and heterochromatin to the Xi. Thus, both chromatin changes on the Xi during B-cell development and the dynamic nature of Xi maintenance in mature B cells predisposes X-linked immunity genes to reactivation [257, 258].

Also, miRNAs can regulate chromatin changes and influence the outcome of T-cell differentiation. Such is the case miR-155 was shown to be highly expressed in dextran sulfate sodium (DSS)-treated mice and to modulate Th17 cells differentiation and function through targeting of Jumonji, AT Rich Interactive Domain 2 (Jarid2). This is a DNA-binding protein that promotes the recruitment of the Polycomb Repressive Complex 2 (PRC2) to chromatin, mediating the transcriptional regulation of IL-22 through regulation of H3K27 methylation and RNA polymerase II transcription complex [178].

Recently, a lncRNA in CD4+ T cells—MEG3—was shown to act as miR-23a sponge, a means by which it modulated gene expression and T-cell differentiation [259]. MEG3 was shown to modulate CD4+ T-cell proliferation, IFN-γ, and TNF-α, T-bet, RORγT, and TIGIT expression. MEG3 overexpression sequestered miR-23 and prevented Th1 and Th17 cell expansion; and attenuated the increase in serum IFN-γ and TNF-α levels in a mouse model.

An example of a lncRNA that codes for miRNAs is that of B-cell integration cluster (BIC), which is differentially expressed during B-cell differentiation [260, 261]. Although its function has not been fully characterized, *BIC* consists of three exons spanning a 13 kb region at chromosome 21q21 and its processed products include miR-155-5p and miR-155-3p. Thus studies related to BIC are essentially focused on these miRNAs, which play a key role in several immune processes, such as hematopoiesis, inflammation, and immune responses [262]. Large-scale cloning studies [263] have allowed to also name BIC as MIR55 host gene or *MIR155HG* (http://www.genenames.org/) and *BIC* transcript as pri-miR-155. BIC is not only highly expressed in antigen receptor-stimulated B and T cells, macrophages, and dendritic cells but also in many types of mature B-cell malignancies, including diffuse large B-cell lymphoma (DLBCL) and chronic lymphocytic leukemia (CLL) [262].

Concluding remarks

Our understanding of immune T- and B-cell differentiation is becoming increasingly more complex and thorough. Initially thought to largely depend on transcription regulation, it was then shown to additionally depend on chromatin modifications and remodeling. Another layer of regulation was unveiled with the recognition of noncoding RNAs as regulators of gene expression and the development of new tools to assess their function. There is a considerable interplay between these layers of gene expression regulation in T- and B-cell lineage differentiation, where both positive and negative regulatory events are crucial. On one hand, transcription factor activation has been associated with the major transitions between the various differentiation stages. Their action can promote local alterations of histone modifications, nucleosome spacing, and DNA methylation through recruitment of histone modification enzymes that complex, with DNA to local sites. Thus changes in epigenetic marks can follow the activity of transcription factors. On the other hand, transcription factors can themselves be regulated by noncoding RNA molecules such as miRNAs and lncRNAs, which in turn can shut down their action and act as brakes in the differentiation process. Other repressive mechanisms based, for example, on histone modifications are of crucial relevance for the maintenance of the undifferentiated stages. They are also the basis for the loss of access to alternative fates, which in turn are irreversibly blocked off at different stages. Ultimately, silencing and activating mechanisms for regulatory genes controlling both T- and B-cell fates are gene- and cell-type specific. Unique epigenetic networks involving chromatin remodeling, histone modifications, and noncoding RNA-mediated molecules act on specific regulatory genes involved in the alternative fates. These, in turn, can remodel chromatin and compete with noncoding RNAs. It remains to be defined which mechanisms act upstream or downstream, a difficult question to mechanistically approach.

The future promises large-scale analyses of these various regulatory components, including at the single-cell level (as technologies continue to improve). The latter will be rather important as to deconstruct concepts that have been built on highly heterogeneous populations. This is particularly the case of T helper cell subsets, often analyzed after highly artificial, so-called polarizing, in vitro conditions. We believe that single-cell technologies will allow unprecedented dissection of T-cell, as well as B-cell, biology directly ex vivo: at the genetic (especially relevant for immunodeficiencies and malignancies), transcriptional and epigenetic levels, thus unveiling the underlying molecular network of the regulatory processes. Importantly, these single-cell technologies will enable groundbreaking analyses in human lymphocytes, thus leading to improved translation of knowledge to human disease. In this regard, we think the epigenetic mechanisms reviewed in this chapter, both chromatin- and RNA-based, will constitute unique opportunities for intervention through innovative chemical and biological design.

References

[1] Kondo M, Weissman IL, Akashi K. Identification of clonogenic common lymphoid progenitors in mouse bone marrow. Cell 1997;91(5):661–72.
[2] Nutt SL, et al. The generation of antibody-secreting plasma cells. Nat Rev Immunol 2015;15(3):160–71.
[3] Fontenot JD, Gavin MA, Rudensky AY. Foxp3 programs the development and function of CD4+CD25+ regulatory T cells. Nat Immunol 2003;4(4):330–6.
[4] Hori S, Nomura T, Sakaguchi S. Control of regulatory T cell development by the transcription factor Foxp3. Science 2003;299(5609):1057–61.

[5] Ivanov II, et al. The orphan nuclear receptor RORgammat directs the differentiation program of proinflammatory IL-17+ T helper cells. Cell 2006;126(6):1121–33.
[6] Szabo SJ, et al. A novel transcription factor, T-bet, directs Th1 lineage commitment. Cell 2000;100 (6):655–69.
[7] Zheng W, Flavell RA. The transcription factor GATA-3 is necessary and sufficient for Th2 cytokine gene expression in CD4 T cells. Cell 1997;89(4):587–96.
[8] Zhu J, Yamane H, Paul WE. Differentiation of effector CD4 T cell populations (*). Annu Rev Immunol 2010;28:445–89.
[9] Sakaguchi S, et al. The plasticity and stability of regulatory T cells. Nat Rev Immunol 2013;13 (6):461–7.
[10] Zan H, Casali P. Epigenetics of peripheral B-cell differentiation and the antibody response. Front Immunol 2015;6:631.
[11] Boller S, Li R, Grosschedl R. Defining B cell chromatin: lessons from EBF1. Trends Genet 2018;34 (4):257–69.
[12] Wu H, et al. Epigenetic regulation in B-cell maturation and its dysregulation in autoimmunity. Cell Mol Immunol 2018;15(7):676–84.
[13] Rodriguez RM, Lopez-Larrea C, Suarez-Alvarez B. Epigenetic dynamics during CD4(+) T cells lineage commitment. Int J Biochem Cell Biol 2015;67:75–85.
[14] Kitagawa Y, Sakaguchi S. Molecular control of regulatory T cell development and function. Curr Opin Immunol 2017;49:64–70.
[15] Bhat J, et al. Stochastics of cellular differentiation explained by epigenetics: the case of T-cell differentiation and functional plasticity. Scand J Immunol 2017;86(4):184–95.
[16] Allan RS, Nutt SL. Deciphering the epigenetic code of T lymphocytes. Immunol Rev 2014;261 (1):50–61.
[17] Gray SM, Kaech SM, Staron MM. The interface between transcriptional and epigenetic control of effector and memory CD8(+) T-cell differentiation. Immunol Rev 2014;261(1):157–68.
[18] Shih HY, et al. Transcriptional and epigenetic networks of helper T and innate lymphoid cells. Immunol Rev 2014;261(1):23–49.
[19] Tripathi SK, Lahesmaa R. Transcriptional and epigenetic regulation of T-helper lineage specification. Immunol Rev 2014;261(1):62–83.
[20] Cedar H, Bergman Y. Epigenetics of haematopoietic cell development. Nat Rev Immunol 2011;11 (7):478–88.
[21] Wilson CB, Rowell E, Sekimata M. Epigenetic control of T-helper-cell differentiation. Nat Rev Immunol 2009;9(2):91–105.
[22] Bird AP. CpG-rich islands and the function of DNA methylation. Nature 1986;321(6067):209–13.
[23] Okano M, et al. DNA methyltransferases Dnmt3a and Dnmt3b are essential for de novo methylation and mammalian development. Cell 1999;99(3):247–57.
[24] Jones, P.A. and G. Liang, Rethinking how DNA methylation patterns are maintained. Nat Rev Genet, 2009. 10(11): p. 805-11.
[25] Smith ZD, Meissner A. DNA methylation: roles in mammalian development. Nat Rev Genet 2013;14(3):204–20.
[26] Attwood JT, Yung RL, Richardson BC. DNA methylation and the regulation of gene transcription. Cell Mol Life Sci 2002;59(2):241–57.
[27] Ambrosi C, Manzo M, Baubec T. Dynamics and context-dependent roles of DNA methylation. J Mol Biol 2017;429(10):1459–75.
[28] Boyes J, Bird A. DNA methylation inhibits transcription indirectly via a methyl-CpG binding protein. Cell 1991;64(6):1123–34.
[29] Jones PL, et al. Methylated DNA and MeCP2 recruit histone deacetylase to repress transcription. Nat Genet 1998;19(2):187–91.
[30] Jones PA. Functions of DNA methylation: islands, start sites, gene bodies and beyond. Nat Rev Genet 2012;13(7):484–92.
[31] Kriaucionis S, Heintz N. The nuclear DNA base 5-hydroxymethylcytosine is present in Purkinje neurons and the brain. Science 2009;324(5929):929–30.

[32] Tahiliani M, et al. Conversion of 5-methylcytosine to 5-hydroxymethylcytosine in mammalian DNA by MLL partner TET1. Science 2009;324(5929):930–5.

[33] Ito S, et al. Tet proteins can convert 5-methylcytosine to 5-formylcytosine and 5-carboxylcytosine. Science 2011;333(6047):1300–3.

[34] Ji H, et al. Comprehensive methylome map of lineage commitment from haematopoietic progenitors. Nature 2010;467(7313):338–42.

[35] Pillay LM, et al. The Hox cofactors Meis1 and Pbx act upstream of gata1 to regulate primitive hematopoiesis. Dev Biol 2010;340(2):306–17.

[36] Molina TJ, et al. Profound block in thymocyte development in mice lacking p56lck. Nature 1992;357 (6374):161–4.

[37] Accomando WP, et al. Quantitative reconstruction of leukocyte subsets using DNA methylation. Genome Biol 2014;15(3):R50.

[38] Schoenborn JR, et al. Comprehensive epigenetic profiling identifies multiple distal regulatory elements directing transcription of the gene encoding interferon-gamma. Nat Immunol 2007;8 (7):732–42.

[39] Lee DU, Agarwal S, Rao A. Th2 lineage commitment and efficient IL-4 production involves extended demethylation of the IL-4 gene. Immunity 2002;16(5):649–60.

[40] Thomas RM, et al. De novo DNA methylation is required to restrict T helper lineage plasticity. J Biol Chem 2012;287(27):22900–9.

[41] Broske AM, et al. DNA methylation protects hematopoietic stem cell multipotency from myeloerythroid restriction. Nat Genet 2009;41(11):1207–15.

[42] Trowbridge JJ, et al. DNA methyltransferase 1 is essential for and uniquely regulates hematopoietic stem and progenitor cells. Cell Stem Cell 2009;5(4):442–9.

[43] Hodges E, et al. Directional DNA methylation changes and complex intermediate states accompany lineage specificity in the adult hematopoietic compartment. Mol Cell 2011;44(1):17–28.

[44] Lee PP, et al. A critical role for Dnmt1 and DNA methylation in T cell development, function, and survival. Immunity 2001;15(5):763–74.

[45] Bird JJ, et al. Helper T cell differentiation is controlled by the cell cycle. Immunity 1998;9(2):229–37.

[46] Tadokoro Y, et al. De novo DNA methyltransferase is essential for self-renewal, but not for differentiation, in hematopoietic stem cells. J Exp Med 2007;204(4):715–22.

[47] Rodriguez RM, et al. Regulation of the transcriptional program by DNA methylation during human alphabeta T-cell development. Nucleic Acids Res 2015;43(2):760–74.

[48] Yu Q, et al. DNA methyltransferase 3a limits the expression of interleukin-13 in T helper 2 cells and allergic airway inflammation. Proc Natl Acad Sci U S A 2012;109(2):541–6.

[49] Ladle BH, et al. De novo DNA methylation by DNA methyltransferase 3a controls early effector CD8 + T-cell fate decisions following activation. Proc Natl Acad Sci U S A 2016;113(38):10631–6.

[50] Lewis JD, et al. Purification, sequence, and cellular localization of a novel chromosomal protein that binds to methylated DNA. Cell 1992;69(6):905–14.

[51] Meehan RR, Lewis JD, Bird AP. Characterization of MeCP2, a vertebrate DNA binding protein with affinity for methylated DNA. Nucleic Acids Res 1992;20(19):5085–92.

[52] Jiang S, et al. MeCP2 reinforces STAT3 signaling and the generation of effector CD4+ T cells by promoting miR-124-mediated suppression of SOCS5. Sci Signal 2014;**7**(316). p. ra25.

[53] Hutchins AS, et al. Gene silencing quantitatively controls the function of a developmental transactivator. Mol Cell 2002;10(1):81–91.

[54] Ichiyama K, et al. The methylcytosine dioxygenase Tet2 promotes DNA demethylation and activation of cytokine gene expression in T cells. Immunity 2015;42(4):613–26.

[55] Tsagaratou A, et al. Dissecting the dynamic changes of 5-hydroxymethylcytosine in T-cell development and differentiation. Proc Natl Acad Sci U S A 2014;111(32). p. E3306-15.

[56] Orlanski S, et al. Tissue-specific DNA demethylation is required for proper B-cell differentiation and function. Proc Natl Acad Sci U S A 2016;113(18):5018–23.

[57] Ooi SK, et al. DNMT3L connects unmethylated lysine 4 of histone H3 to de novo methylation of DNA. Nature 2007;448(7154):714–7.

[58] Otani J, et al. Structural basis for recognition of H3K4 methylation status by the DNA methyltransferase 3A ATRX-DNMT3-DNMT3L domain. EMBO Rep 2009;10(11):1235–41.

[59] Noh KM, et al. Engineering of a Histone-Recognition Domain in Dnmt3a Alters the Epigenetic Landscape and Phenotypic Features of Mouse ESCs. Mol Cell 2015;59(1):89–103.
[60] Rothbart SB, et al. Association of UHRF1 with methylated H3K9 directs the maintenance of DNA methylation. Nat Struct Mol Biol 2012;19(11):1155–60.
[61] Liu X, et al. UHRF1 targets DNMT1 for DNA methylation through cooperative binding of hemi-methylated DNA and methylated H3K9. Nat Commun 2013;4:1563.
[62] Dhayalan A, et al. The Dnmt3a PWWP domain reads histone 3 lysine 36 trimethylation and guides DNA methylation. J Biol Chem 2010;285(34):26114–20.
[63] Baubec T, et al. Genomic profiling of DNA methyltransferases reveals a role for DNMT3B in genic methylation. Nature 2015;520(7546):243–7.
[64] Rondelet G, et al. Structural basis for recognition of histone H3K36me3 nucleosome by human de novo DNA methyltransferases 3A and 3B. J Struct Biol 2016;194(3):357–67.
[65] Jenuwein T, Allis CD. Translating the histone code. Science 2001;293(5532):1074–80.
[66] Suganuma T, Workman JL. Crosstalk among Histone Modifications. Cell 2008;135(4):604–7.
[67] Barski A, et al. High-resolution profiling of histone methylations in the human genome. Cell 2007;129(4):823–37.
[68] Wang Z, et al. Combinatorial patterns of histone acetylations and methylations in the human genome. Nat Genet 2008;40(7):897–903.
[69] Dose, M., et al., beta-Catenin induces T-cell transformation by promoting genomic instability. Proc Natl Acad Sci U S A, 2014. 111(1): p. 391-6.
[70] Zhang JA, et al. Dynamic transformations of genome-wide epigenetic marking and transcriptional control establish T cell identity. Cell 2012;149(2):467–82.
[71] Avni O, et al. T(H) cell differentiation is accompanied by dynamic changes in histone acetylation of cytokine genes. Nat Immunol 2002;3(7):643–51.
[72] Denton AE, et al. Differentiation-dependent functional and epigenetic landscapes for cytokine genes in virus-specific CD8+ T cells. Proc Natl Acad Sci U S A 2011;108(37):15306–11.
[73] Ren Q, Gorovsky MA. Histone H2A.Z acetylation modulates an essential charge patch. Mol Cell 2001;7(6):1329–35.
[74] Heintzman ND, et al. Histone modifications at human enhancers reflect global cell-type-specific gene expression. Nature 2009;459(7243):108–12.
[75] Cairns BR. The logic of chromatin architecture and remodelling at promoters. Nature 2009;461 (7261):193–8.
[76] Talbert PB, Henikoff S. Histone variants—ancient wrap artists of the epigenome. Nat Rev Mol Cell Biol 2010;11(4):264–75.
[77] Agarwal S, Rao A. Modulation of chromatin structure regulates cytokine gene expression during T cell differentiation. Immunity 1998;9(6):765–75.
[78] Koyanagi M, et al. EZH2 and histone 3 trimethyl lysine 27 associated with Il4 and Il13 gene silencing in Th1 cells. J Biol Chem 2005;280(36):31470–7.
[79] Roh TY, et al. The genomic landscape of histone modifications in human T cells. Proc Natl Acad Sci U S A 2006;103(43):15782–7.
[80] Wei G, et al. Global mapping of H3K4me3 and H3K27me3 reveals specificity and plasticity in lineage fate determination of differentiating CD4+ T cells. Immunity 2009;30(1):155–67.
[81] Araki Y, et al. Genome-wide analysis of histone methylation reveals chromatin state-based regulation of gene transcription and function of memory CD8+ T cells. Immunity 2009;30(6):912–25.
[82] Tumes DJ, et al. The polycomb protein Ezh2 regulates differentiation and plasticity of CD4(+) T helper type 1 and type 2 cells. Immunity 2013;39(5):819–32.
[83] Hawkins RD, et al. Global chromatin state analysis reveals lineage-specific enhancers during the initiation of human T helper 1 and T helper 2 cell polarization. Immunity 2013;38(6):1271–84.
[84] Hirota K, et al. Fate mapping of IL-17-producing T cells in inflammatory responses. Nat Immunol 2011;12(3):255–63.
[85] Lee YK, et al. Late developmental plasticity in the T helper 17 lineage. Immunity 2009;30(1):92–107.
[86] Dominguez-Villar M, Baecher-Allan CM, Hafler DA. Identification of T helper type 1-like, Foxp3+ regulatory T cells in human autoimmune disease. Nat Med 2011;17(6):673–5.

[87] Oldenhove G, et al. Decrease of Foxp3+ Treg cell number and acquisition of effector cell phenotype during lethal infection. Immunity 2009;31(5):772–86.

[88] Schmolka N, et al. Epigenetic and transcriptional signatures of stable versus plastic differentiation of proinflammatory gammadelta T cell subsets. Nat Immunol 2013;14(10):1093–100.

[89] Allan RS, et al. An epigenetic silencing pathway controlling T helper 2 cell lineage commitment. Nature 2012;487(7406):249–53.

[90] Zhang Y, et al. The polycomb repressive complex 2 governs life and death of peripheral T cells. Blood 2014;124(5):737–49.

[91] Yang XP, et al. EZH2 is crucial for both differentiation of regulatory T cells and T effector cell expansion. Sci Rep 2015;5:10643.

[92] DuPage M, et al. The chromatin-modifying enzyme Ezh2 is critical for the maintenance of regulatory T cell identity after activation. Immunity 2015;42(2):227–38.

[93] Sarmento OF, et al. The Role of the Histone Methyltransferase Enhancer of Zeste Homolog 2 (EZH2) in the Pathobiological Mechanisms Underlying Inflammatory Bowel Disease (IBD). J Biol Chem 2017;292(2):706–22.

[94] Ikawa T, et al. Conversion of T cells to B cells by inactivation of polycomb-mediated epigenetic suppression of the B-lineage program. Genes Dev 2016;30(22):2475–85.

[95] Aloia L, Di Stefano B, Di Croce L. Polycomb complexes in stem cells and embryonic development. Development 2013;140(12):2525–34.

[96] Ikawa T, et al. Long-term cultured E2A-deficient hematopoietic progenitor cells are pluripotent. Immunity 2004;20(3):349–60.

[97] Miyazaki M, et al. Polycomb group gene mel-18 regulates early T progenitor expansion by maintaining the expression of Hes-1, a target of the Notch pathway. J Immunol 2005;174(5):2507–16.

[98] Cales C, et al. Inactivation of the polycomb group protein Ring1B unveils an antiproliferative role in hematopoietic cell expansion and cooperation with tumorigenesis associated with Ink4a deletion. Mol Cell Biol 2008;28(3):1018–28.

[99] Oguro H, et al. Poised lineage specification in multipotential hematopoietic stem and progenitor cells by the polycomb protein Bmi1. Cell Stem Cell 2010;6(3):279–86.

[100] Yamashita M, et al. Crucial role of MLL for the maintenance of memory T helper type 2 cell responses. Immunity 2006;24(5):611–22.

[101] Zhao DM, Xue HH. MLL4 keeps Foxp3 in the loop. Nat Immunol 2017;18(9):957–8.

[102] Placek K, et al. MLL4 prepares the enhancer landscape for Foxp3 induction via chromatin looping. Nat Immunol 2017;18(9):1035–45.

[103] Mele DA, et al. BET bromodomain inhibition suppresses TH17-mediated pathology. J Exp Med 2013;210(11):2181–90.

[104] Kimura M, et al. Regulation of Th2 cell differentiation by mel-18, a mammalian polycomb group gene. Immunity 2001;15(2):275–87.

[105] Azagra A, et al. In vivo conditional deletion of HDAC7 reveals its requirement to establish proper B lymphocyte identity and development. J Exp Med 2016;213(12):2591–601.

[106] Valapour M, et al. Histone deacetylation inhibits IL4 gene expression in T cells. J Allergy Clin Immunol 2002;109(2):238–45.

[107] Beier UH, et al. Histone deacetylases 6 and 9 and sirtuin-1 control Foxp3+ regulatory T cell function through shared and isoform-specific mechanisms. Sci Signal 2012;5(229). p. ra45.

[108] Mesin L, Ersching J, Victora GD. Germinal center B cell dynamics. Immunity 2016;45(3):471–82.

[109] Ortega-Molina A, et al. The histone lysine methyltransferase KMT2D sustains a gene expression program that represses B cell lymphoma development. Nat Med 2015;21(10):1199–208.

[110] Zhang J, et al. Disruption of KMT2D perturbs germinal center B cell development and promotes lymphomagenesis. Nat Med 2015;21(10):1190–8.

[111] Zhang J, et al. The CREBBP acetyltransferase is a Haploinsufficient tumor suppressor in B-cell lymphoma. Cancer Discov 2017;7(3):322–37.

[112] Jiang Y, et al. CREBBP inactivation promotes the development of HDAC3-dependent lymphomas. Cancer Discov 2017;7(1):38–53.

[113] Hatzi K, et al. Histone demethylase LSD1 is required for germinal center formation and BCL6-driven lymphomagenesis. Nat Immunol 2019;20(1):86–96.

[114] Shi Y, et al. Histone demethylation mediated by the nuclear amine oxidase homolog LSD1. Cell 2004;119(7):941–53.
[115] Kerenyi MA, et al. Histone demethylase Lsd1 represses hematopoietic stem and progenitor cell signatures during blood cell maturation. Elife 2013;2:e00633.
[116] Klemm SL, Shipony Z, Greenleaf WJ. Chromatin accessibility and the regulatory epigenome. Nat Rev Genet 2019;20(4):207–20.
[117] Cusanovich DA, et al. Multiplex single cell profiling of chromatin accessibility by combinatorial cellular indexing. Science 2015;348(6237):910–4.
[118] Cusanovich DA, et al. The cis-regulatory dynamics of embryonic development at single-cell resolution. Nature 2018;555(7697):538–42.
[119] Cusanovich DA, et al. A single-cell atlas of in vivo mammalian chromatin accessibility. Cell 2018;174 (5):1309–24. e18.
[120] Scott-Browne JP, et al. Dynamic changes in chromatin accessibility occur in CD8(+) T cells responding to viral infection. Immunity 2016;45(6):1327–40.
[121] Boller S, et al. Pioneering activity of the C-terminal domain of EBF1 shapes the chromatin landscape for B cell programming. Immunity 2016;44(3):527–41.
[122] Johnson JL, et al. Lineage-determining transcription factor TCF-1 initiates the epigenetic identity of T cells. Immunity 2018;48(2):243–57. e10.
[123] Yoshida H, et al. The cis-regulatory atlas of the mouse immune system. Cell 2019;176(4):897–912 e20.
[124] Lau CM, et al. Epigenetic control of innate and adaptive immune memory. Nat Immunol 2018;19 (9):963–72.
[125] Dai R, Ahmed SA. MicroRNA, a new paradigm for understanding immunoregulation, inflammation, and autoimmune diseases. Transl Res 2011;157(4):163–79.
[126] Baltimore D, et al. MicroRNAs: new regulators of immune cell development and function. Nat Immunol 2008;9(8):839–45.
[127] O'Neill LA, Sheedy FJ, McCoy CE. MicroRNAs: the fine-tuners of Toll-like receptor signalling. Nat Rev Immunol 2011;11(3):163–75.
[128] Baumjohann D, Ansel KM. MicroRNA-mediated regulation of T helper cell differentiation and plasticity. Nat Rev Immunol 2013;13(9):666–78.
[129] Casero D, et al. Long non-coding RNA profiling of human lymphoid progenitor cells reveals transcriptional divergence of B cell and T cell lineages. Nat Immunol 2015;16(12):1282–91.
[130] Aalto AP, Pasquinelli AE. Small non-coding RNAs mount a silent revolution in gene expression. Curr Opin Cell Biol 2012;24(3):333–40.
[131] Jinek M, Doudna JA. A three-dimensional view of the molecular machinery of RNA interference. Nature 2009;457(7228):405–12.
[132] Zeng Y, Cullen BR. Structural requirements for pre-microRNA binding and nuclear export by Exportin 5. Nucleic Acids Res 2004;32(16):4776–85.
[133] Meister G. Argonaute proteins: functional insights and emerging roles. Nat Rev Genet 2013;14 (7):447–59.
[134] Khvorova A, Reynolds A, Jayasena SD. Functional siRNAs and miRNAs exhibit strand bias. Cell 2003;115(2):209–16.
[135] Moazed D. Small RNAs in transcriptional gene silencing and genome defence. Nature 2009;457 (7228):413–20.
[136] Creamer KM, Partridge JF. RITS-connecting transcription, RNA interference, and heterochromatin assembly in fission yeast. Wiley Interdiscip Rev RNA 2011;2(5):632–46.
[137] Guo H, et al. Mammalian microRNAs predominantly act to decrease target mRNA levels. Nature 2010;466(7308):835–40.
[138] Bernstein E, et al. Dicer is essential for mouse development. Nat Genet 2003;35(3):215–7.
[139] Steiner DF, et al. MicroRNA-29 regulates T-box transcription factors and interferon-gamma production in helper T cells. Immunity 2011;35(2):169–81.
[140] Chong MM, et al. The RNAseIII enzyme Drosha is critical in T cells for preventing lethal inflammatory disease. J Exp Med 2008;205(9):2005–17.

[141] Muljo SA, et al. Aberrant T cell differentiation in the absence of Dicer. J Exp Med 2005;202(2):261–9.
[142] Zhang N, Bevan MJ. Dicer controls CD8+ T-cell activation, migration, and survival. Proc Natl Acad Sci U S A 2010;107(50):21629–34.
[143] Trifari S, et al. MicroRNA-directed program of cytotoxic CD8+ T-cell differentiation. Proc Natl Acad Sci U S A 2013;110(46):18608–13.
[144] Pua HH, et al. MicroRNAs 24 and 27 suppress allergic inflammation and target a network of regulators of T helper 2 cell-associated cytokine production. Immunity 2016;44(4):821–32.
[145] Baumjohann D, et al. The microRNA cluster miR-17∼92 promotes TFH cell differentiation and represses subset-inappropriate gene expression. Nat Immunol 2013;14(8):840–8.
[146] Zhou L, et al. MicroRNAs are key regulators controlling iNKT and regulatory T-cell development and function. Cell Mol Immunol 2011;8(5):380–7.
[147] Cobb BS, et al. A role for Dicer in immune regulation. J Exp Med 2006;203(11):2519–27.
[148] Fedeli M, et al. miR-17∼92 family clusters control iNKT cell ontogenesis via modulation of TGF-beta signaling. Proc Natl Acad Sci U S A 2016;113(51):E8286–95.
[149] Koralov SB, et al. Dicer ablation affects antibody diversity and cell survival in the B lymphocyte lineage. Cell 2008;132(5):860–74.
[150] Xu S, et al. The RNase III enzyme Dicer is essential for germinal center B-cell formation. Blood 2012;119(3):767–76.
[151] Belver L, de Yebenes VG, Ramiro AR. MicroRNAs prevent the generation of autoreactive antibodies. Immunity 2010;33(5):713–22.
[152] Liang Y, Pan HF, Ye DQ. microRNAs function in CD8+T cell biology. J Leukoc Biol 2015;97(3):487–97.
[153] Jeker LT, Bluestone JA. MicroRNA regulation of T-cell differentiation and function. Immunol Rev 2013;253(1):65–81.
[154] Smith KM, et al. miR-29ab1 deficiency identifies a negative feedback loop controlling Th1 bias that is dysregulated in multiple sclerosis. J Immunol 2012;189(4):1567–76.
[155] Ma F, et al. The microRNA miR-29 controls innate and adaptive immune responses to intracellular bacterial infection by targeting interferon-gamma. Nat Immunol 2011;12(9):861–9.
[156] Guan H, et al. Inverse correlation of expression of microRNA-140-5p with progression of multiple sclerosis and differentiation of encephalitogenic T helper type 1 cells. Immunology 2016;147(4):488–98.
[157] Mohnle P, et al. MicroRNA-146a controls Th1-cell differentiation of human CD4+ T lymphocytes by targeting PRKCepsilon. Eur J Immunol 2015;45(1):260–72.
[158] Wu W, et al. miR-10a inhibits dendritic cell activation and Th1/Th17 cell immune responses in IBD. Gut 2015;64(11):1755–64.
[159] Huffaker TB, et al. Epistasis between microRNAs 155 and 146a during T cell-mediated antitumor immunity. Cell Rep 2012;2(6):1697–709.
[160] Jiang S, et al. Molecular dissection of the miR-17-92 cluster's critical dual roles in promoting Th1 responses and preventing inducible Treg differentiation. Blood 2011;118(20):5487–97.
[161] Cho S, et al. A novel miR-24-TCF1 axis in modulating effector T cell responses. J Immunol 2017;198(10):3919–26.
[162] Wu F, et al. MicroRNAs are differentially expressed in ulcerative colitis and alter expression of macrophage inflammatory peptide-2 alpha. Gastroenterology 2008;135(5):1624–35. e24.
[163] Murata K, et al. Comprehensive microRNA analysis identifies miR-24 and miR-125a-5p as plasma biomarkers for rheumatoid arthritis. PLoS One 2013;8(7)e69118.
[164] Cho S, et al. miR-23∼27∼24 clusters control effector T cell differentiation and function. J Exp Med 2016;213(2):235–49.
[165] Xin Q, et al. miR-155 deficiency ameliorates autoimmune inflammation of systemic lupus Erythematosus by targeting S1pr1 in Faslpr/lpr mice. J Immunol 2015;194(11):5437–45.
[166] Okoye IS, et al. Transcriptomics identified a critical role for Th2 cell-intrinsic miR-155 in mediating allergy and antihelminth immunity. Proc Natl Acad Sci U S A 2014;111(30):E3081–90.
[167] Simpson LJ, et al. A microRNA upregulated in asthma airway T cells promotes TH2 cytokine production. Nat Immunol 2014;15(12):1162–70.

[168] Kastle M, et al. microRNA cluster 106a∼363 is involved in T helper 17 cell differentiation. Immunology 2017;152(3):402–13.
[169] Montoya MM, et al. A distinct inhibitory function for miR-18a in Th17 cell differentiation. J Immunol 2017;199(2):559–69.
[170] Li B, et al. miR-146a modulates autoreactive Th17 cell differentiation and regulates organ-specific autoimmunity. J Clin Invest 2017;127(10):3702–16.
[171] Du C, et al. MicroRNA miR-326 regulates TH-17 differentiation and is associated with the pathogenesis of multiple sclerosis. Nat Immunol 2009;10(12):1252–9.
[172] Satoorian T, et al. MicroRNA223 promotes pathogenic T-cell development and autoimmune inflammation in central nervous system in mice. Immunology 2016;148(4):326–38.
[173] Qu X, et al. MiR-384 regulates the Th17/Treg ratio during experimental autoimmune encephalomyelitis pathogenesis. Front Cell Neurosci 2017;11:88.
[174] Liu Q, et al. MicroRNA-590 promotes pathogenic Th17 cell differentiation through targeting Tob1 and is associated with multiple sclerosis. Biochem Biophys Res Commun 2017;493(2):901–8.
[175] Schulze-Topphoff U, et al. Tob1 plays a critical role in the activation of encephalitogenic T cells in CNS autoimmunity. J Exp Med 2013;210(7):1301–9.
[176] Wu R, et al. MicroRNA-448 promotes multiple sclerosis development through induction of Th17 response through targeting protein tyrosine phosphatase non-receptor type 2 (PTPN2). Biochem Biophys Res Commun 2017;486(3):759–66.
[177] Escobar TM, et al. miR-155 activates cytokine gene expression in Th17 cells by regulating the DNA-binding protein Jarid2 to relieve polycomb-mediated repression. Immunity 2014;40(6):865–79.
[178] Xu M, et al. MiR-155 contributes to Th17 cells differentiation in dextran sulfate sodium (DSS)-induced colitis mice via Jarid2. Biochem Biophys Res Commun 2017;488(1):6–14.
[179] Liu L, et al. Elevated expression of microRNA-873 facilitates Th17 differentiation by targeting forkhead box O1 (Foxo1) in the pathogenesis of systemic lupus erythematosus. Biochem Biophys Res Commun 2017;492(3):453–60.
[180] Yang X, et al. MicroRNA-425 facilitates pathogenic Th17 cell differentiation by targeting forkhead box O1 (Foxo1) and is associated with inflammatory bowel disease. Biochem Biophys Res Commun 2018;496(2):352–8.
[181] Zhang L, et al. MicroRNA-31 negatively regulates peripherally derived regulatory T-cell generation by repressing retinoic acid-inducible protein 3. Nat Commun 2015;6:7639.
[182] Cruz LO, et al. Excessive expression of miR-27 impairs Treg-mediated immunological tolerance. J Clin Invest 2017;127(2):530–42.
[183] Pan W, et al. MiR-125a targets effector programs to stabilize Treg-mediated immune homeostasis. Nat Commun 2015;6:7096.
[184] Sanchez-Diaz R, et al. Thymus-derived regulatory T cell development is regulated by C-type lectin-mediated BIC/MicroRNA 155 expression. Mol Cell Biol 2017;**37**(9).
[185] Pratama A, et al. MicroRNA-146a regulates ICOS-ICOSL signalling to limit accumulation of T follicular helper cells and germinal centres. Nat Commun 2015;6:6436.
[186] Hu R, et al. miR-155 promotes T follicular helper cell accumulation during chronic, low-grade inflammation. Immunity 2014;41(4):605–19.
[187] Liu WH, et al. A miR-155-Peli1-c-Rel pathway controls the generation and function of T follicular helper cells. J Exp Med 2016;213(9):1901–19.
[188] Wu T, et al. Cutting edge: miR-17-92 is required for both CD4 Th1 and T follicular helper cell responses during viral infection. J Immunol 2015;195(6):2515–9.
[189] Kang SG, et al. MicroRNAs of the miR-17∼92 family are critical regulators of T(FH) differentiation. Nat Immunol 2013;14(8):849–57.
[190] Wells AC, et al. Modulation of let-7 miRNAs controls the differentiation of effector CD8 T cells. Elife 2017;6.
[191] Wang S, et al. MicroRNA-146a feedback suppresses T cell immune function by targeting Stat1 in patients with chronic hepatitis B. J Immunol 2013;191(1):293–301.
[192] Yu T, et al. MicroRNA-491 regulates the proliferation and apoptosis of CD8(+) T cells. Sci Rep 2016;6:30923.

[193] Lin R, et al. Targeting miR-23a in CD8+ cytotoxic T lymphocytes prevents tumor-dependent immunosuppression. J Clin Invest 2014;124(12):5352–67.

[194] Tarasenko T, et al. T cell-specific deletion of the inositol phosphatase SHIP reveals its role in regulating Th1/Th2 and cytotoxic responses. Proc Natl Acad Sci U S A 2007;104(27):11382–7.

[195] Dudda JC, et al. MicroRNA-155 is required for effector CD8+ T cell responses to virus infection and cancer. Immunity 2013;38(4):742–53.

[196] Gracias DT, et al. The microRNA miR-155 controls CD8(+) T cell responses by regulating interferon signaling. Nat Immunol 2013;14(6):593–602.

[197] Ji Y, et al. miR-155 augments CD8+ T-cell antitumor activity in lymphoreplete hosts by enhancing responsiveness to homeostatic gammac cytokines. Proc Natl Acad Sci U S A 2015;112(2):476–81.

[198] Chen Z, et al. miR-150 regulates memory CD8 T cell differentiation via c-Myb. Cell Rep 2017;20 (11):2584–97.

[199] Gautam S, et al. The transcription factor c-Myb regulates CD8(+) T cell stemness and antitumor immunity. Nat Immunol 2019;20(3):337–49.

[200] Wu T, et al. Temporal expression of microRNA cluster miR-17-92 regulates effector and memory CD8+ T-cell differentiation. Proc Natl Acad Sci U S A 2012;109(25):9965–70.

[201] Pobezinsky LA, et al. Let-7 microRNAs target the lineage-specific transcription factor PLZF to regulate terminal NKT cell differentiation and effector function. Nat Immunol 2015;16(5):517–24.

[202] Burocchi A, et al. Regulated expression of miR-155 is required for iNKT cell development. Front Immunol 2015;6:140.

[203] Zheng Q, Zhou L, Mi QS. MicroRNA miR-150 is involved in Valpha14 invariant NKT cell development and function. J Immunol 2012;188(5):2118–26.

[204] Henao-Mejia J, et al. The microRNA miR-181 is a critical cellular metabolic rheostat essential for NKT cell ontogenesis and lymphocyte development and homeostasis. Immunity 2013;38(5):984–97.

[205] Blume J, et al. Overexpression of Valpha14Jalpha18 TCR promotes development of iNKT cells in the absence of miR-181a/b-1. Immunol Cell Biol 2016;94(8):741–6.

[206] Di Pietro C, et al. MicroRNA-133b regulation of Th-POK expression and dendritic cell signals affect NKT17 cell differentiation in the thymus. J Immunol 2016;197(8):3271–80.

[207] Schmolka N, et al. MicroRNA-146a controls functional plasticity in gammadelta T cells by targeting NOD1. Sci Immunol 2018;**3**(23).

[208] Amado T, et al. Cross-regulation between cytokine and microRNA pathways in T cells. Eur J Immunol 2015;45(6):1584–95.

[209] Yang L, et al. miR-146a controls the resolution of T cell responses in mice. J Exp Med 2012;209 (9):1655–70.

[210] Lu LF, et al. Function of miR-146a in controlling Treg cell-mediated regulation of Th1 responses. Cell 2010;142(6):914–29.

[211] Takahashi H, et al. TGF-beta and retinoic acid induce the microRNA miR-10a, which targets Bcl-6 and constrains the plasticity of helper T cells. Nat Immunol 2012;13(6):587–95.

[212] Ademokun A, Turner M. Regulation of B-cell differentiation by microRNAs and RNA-binding proteins. Biochem Soc Trans 2008;36(Pt 6):1191–3.

[213] Chen CZ, et al. MicroRNAs modulate hematopoietic lineage differentiation. Science 2004;303 (5654):83–6.

[214] Ventura A, et al. Targeted deletion reveals essential and overlapping functions of the miR-17 through 92 family of miRNA clusters. Cell 2008;132(5):875–86.

[215] Rao DS, et al. MicroRNA-34a perturbs B lymphocyte development by repressing the forkhead box transcription factor Foxp1. Immunity 2010;33(1):48–59.

[216] Chaudhuri AA, et al. Oncomir miR-125b regulates hematopoiesis by targeting the gene Lin28A. Proc Natl Acad Sci U S A 2012;109(11):4233–8.

[217] Gururajan M, et al. MicroRNA 125b inhibition of B cell differentiation in germinal centers. Int Immunol 2010;22(7):583–92.

[218] Puissegur MP, et al. B-cell regulator of immunoglobulin heavy-chain transcription (Bright)/ARID3a is a direct target of the oncomir microRNA-125b in progenitor B-cells. Leukemia 2012;26 (10):2224–32.

[219] Xiao C, et al. MiR-150 controls B cell differentiation by targeting the transcription factor c-Myb. Cell 2007;131(1):146–59.
[220] Mehta A, et al. The microRNA-212/132 cluster regulates B cell development by targeting Sox4. J Exp Med 2015;212(10):1679–92.
[221] Li QJ, et al. miR-181a is an intrinsic modulator of T cell sensitivity and selection. Cell 2007;129 (1):147–61.
[222] Ebert PJ, et al. An endogenous positively selecting peptide enhances mature T cell responses and becomes an autoantigen in the absence of microRNA miR-181a. Nat Immunol 2009;10(11):1162–9.
[223] Fragoso R, et al. Modulating the strength and threshold of NOTCH oncogenic signals by mir-181a-1/b-1. PLoS Genet 2012;8(8). e1002855.
[224] Tanzer A, Stadler PF. Molecular evolution of a microRNA cluster. J Mol Biol 2004;339(2):327–35.
[225] Lai M, et al. Regulation of B-cell development and tolerance by different members of the miR-17 ~92 family microRNAs. Nat Commun 2016;7:12207.
[226] Georgantas 3rd RW, et al. CD34+ hematopoietic stem-progenitor cell microRNA expression and function: a circuit diagram of differentiation control. Proc Natl Acad Sci U S A 2007;104(8):2750–5.
[227] Crotty S. Follicular helper CD4 T cells (TFH). Annu Rev Immunol 2011;29:621–63.
[228] Iyer MK, et al. The landscape of long noncoding RNAs in the human transcriptome. Nat Genet 2015;47(3):199–208.
[229] Derrien T, et al. The GENCODE v7 catalog of human long noncoding RNAs: analysis of their gene structure, evolution, and expression. Genome Res 2012;22(9):1775–89.
[230] Lee JT. Epigenetic regulation by long noncoding RNAs. Science 2012;338(6113):1435–9.
[231] Cabili MN, et al. Integrative annotation of human large intergenic noncoding RNAs reveals global properties and specific subclasses. Genes Dev 2011;25(18):1915–27.
[232] Ulitsky I. Evolution to the rescue: using comparative genomics to understand long non-coding RNAs. Nat Rev Genet 2016;17(10):601–14.
[233] Sauvageau M, et al. Multiple knockout mouse models reveal lincRNAs are required for life and brain development. Elife 2013;2. e01749.
[234] Zhu S, et al. Genome-scale deletion screening of human long non-coding RNAs using a paired-guide RNA CRISPR-Cas9 library. Nat Biotechnol 2016;34(12):1279–86.
[235] Marchese FP, Raimondi I, Huarte M. The multidimensional mechanisms of long noncoding RNA function. Genome Biol 2017;18(1):206.
[236] Carpenter S, et al. A long noncoding RNA mediates both activation and repression of immune response genes. Science 2013;341(6147):789–92.
[237] Rapicavoli NA, et al. A mammalian pseudogene lncRNA at the interface of inflammation and anti-inflammatory therapeutics. Elife 2013;2. e00762.
[238] Hu G, et al. Expression and regulation of intergenic long noncoding RNAs during T cell development and differentiation. Nat Immunol 2013;14(11):1190–8.
[239] Spurlock 3rd CF, et al. Expression and functions of long noncoding RNAs during human T helper cell differentiation. Nat Commun 2015;6:6932.
[240] Ranzani V, et al. The long intergenic noncoding RNA landscape of human lymphocytes highlights the regulation of T cell differentiation by linc-MAF-4. Nat Immunol 2015;16(3):318–25.
[241] Collier SP, et al. Cutting edge: influence of Tmevpg1, a long intergenic noncoding RNA, on the expression of Ifng by Th1 cells. J Immunol 2012;189(5):2084–8.
[242] Gomez JA, et al. The NeST long ncRNA controls microbial susceptibility and epigenetic activation of the interferon-gamma locus. Cell 2013;152(4):743–54.
[243] Huang W, et al. DDX5 and its associated lncRNA Rmrp modulate TH17 cell effector functions. Nature 2015;528(7583):517–22.
[244] Makitie O, Kaitila I, Savilahti E. Susceptibility to infections and in vitro immune functions in cartilage-hair hypoplasia. Eur J Pediatr 1998;157(10):816–20.
[245] Bonafe L, et al. Evolutionary comparison provides evidence for pathogenicity of RMRP mutations. PLoS Genet 2005;1(4)e47.
[246] Masoumi F, et al. Malat1 long noncoding RNA regulates inflammation and leukocyte differentiation in experimental autoimmune encephalomyelitis. J Neuroimmunol 2019;328:50–9.

[247] Chen YG, Satpathy AT, Chang HY. Gene regulation in the immune system by long noncoding RNAs. Nat Immunol 2017;18(9):962–72.
[248] Brazao TF, et al. Long noncoding RNAs in B-cell development and activation. Blood 2016;128(7): e10–9.
[249] Winkle M, et al. Emerging roles for long noncoding RNAs in B-cell development and malignancy. Crit Rev Oncol Hematol 2017;120:77–85.
[250] Tayari MM, et al. Long noncoding RNA expression profiling in normal B-cell subsets and Hodgkin lymphoma reveals Hodgkin and Reed-Sternberg cell-specific long noncoding RNAs. Am J Pathol 2016;186(9):2462–72.
[251] Petri A, et al. Long noncoding RNA expression during human B-cell development. PLoS One 2015;10(9). e0138236.
[252] Herzog S, Reth M, Jumaa H. Regulation of B-cell proliferation and differentiation by pre-B-cell receptor signalling. Nat Rev Immunol 2009;9(3):195–205.
[253] Ellis BC, Graham LD, Molloy PL. CRNDE, a long non-coding RNA responsive to insulin/IGF signaling, regulates genes involved in central metabolism. Biochim Biophys Acta 2014;1843(2):372–86.
[254] Nutt SL, et al. Commitment to the B-lymphoid lineage depends on the transcription factor Pax5. Nature 1999;401(6753):556–62.
[255] Verma-Gaur J, et al. Noncoding transcription within the Igh distal V(H) region at PAIR elements affects the 3D structure of the Igh locus in pro-B cells. Proc Natl Acad Sci U S A 2012;109 (42):17004–9.
[256] Liu H, et al. Yin Yang 1 is a critical regulator of B-cell development. Genes Dev 2007;21 (10):1179–89.
[257] Syrett CM, et al. Loss of Xist RNA from the inactive X during B cell development is restored in a dynamic YY1-dependent two-step process in activated B cells. PLoS Genet 2017;13(10). e1007050.
[258] Wang J, et al. Unusual maintenance of X chromosome inactivation predisposes female lymphocytes for increased expression from the inactive X. Proc Natl Acad Sci U S A 2016;113(14):E2029–38.
[259] Wang J, et al. MEG3 modulates TIGIT expression and CD4 + T cell activation through absorbing miR-23a. Mol Cell Biochem 2019;454(1–2):67–76.
[260] Eis PS, et al. Accumulation of miR-155 and BIC RNA in human B cell lymphomas. Proc Natl Acad Sci U S A 2005;102(10):3627–32.
[261] Tam W. Identification and characterization of human BIC, a gene on chromosome 21 that encodes a noncoding RNA. Gene 2001;274(1–2):157–67.
[262] Landgraf P, et al. A mammalian microRNA expression atlas based on small RNA library sequencing. Cell 2007;129(7):1401–14.
[263] Mullighan CG, Downing JR. Global genomic characterization of acute lymphoblastic leukemia. Semin Hematol 2009;46(1):3–15.

Further reading

Dispirito JR, Shen H. Histone acetylation at the single-cell level: a marker of memory CD8+ T cell differentiation and functionality. J Immunol 2010;184(9):4631–6.
Fann M, et al. Histone acetylation is associated with differential gene expression in the rapid and robust memory CD8(+) T-cell response. Blood 2006;108(10):3363–70.
Northrop JK, et al. Epigenetic remodeling of the IL-2 and IFN-gamma loci in memory CD8 T cells is influenced by CD4 T cells. J Immunol 2006;177(2):1062–9.
Northrop JK, Wells AD, Shen H. Cutting edge: chromatin remodeling as a molecular basis for the enhanced functionality of memory CD8 T cells. J Immunol 2008;181(2):865–8.
Roh TY, Cuddapah S, Zhao K. Active chromatin domains are defined by acetylation islands revealed by genome-wide mapping. Genes Dev 2005;19(5):542–52.

CHAPTER 5

Epigenetics mechanisms driving immune memory cell differentiation and function

Stephen J. Turner[a,b], **Jasmine Li**[a], **Brendan E. Russ**[a]
[a]Department of Microbiology, Biomedical Discovery Institute, Monash University, Clayton, VIC, Australia
[b]Department of Microbiology and Immunology, The Peter Doherty Institute for Infection and Immunity, The University of Melbourne, Parkville, VIC, Australia

Contents

Introduction 117
Functional heterogeneity within the memory T-cell pool 119
Histone methylation and pattern, acquisition of function 120
DNA methylation and its role in regulating gene transcription 122
  Making sense of the "junk DNA": Noncoding regulatory elements work via chromatin folding 122
Epigenetic regulation in the acquisition of lineage-specific T-cell function 123
Epigenetic regulation in the acquisition of CD8 T-cell function 124
Epigenetic mapping of the differentiation pathway that leads to T-cell memory 126
The role of $CD8^+$ T-cell-specific transcription factors in chromatin remodeling and acquisition of function 127
Active regulation of chromatin state is a key factor in $CD8^+$ T-cell effector vs memory fate decisions 129
Conclusion 132
References 133

Introduction

T cells form a major component of the mammalian cellular adaptive immune system and can be divided into two major lineages, namely $CD4^+$ T-helper lymphocytes (T_H) and $CD8^+$ killer T cells. Each lineage has distinct roles in pathogen control and protection from subsequent infection. $CD4^+$ T cells promote effective immunity by producing key cytokines that promote effective antibody and cellular responses upon infection. $CD4^+$ T_H cells can differentiate into different subtypes characterized by expression of different effector molecules and thus have different roles in immunity to infection. $CD8^+$ T cells are considered the "hit-men" of the immune system, locating and destroying virus-infected cells, thus limiting and contributing to the eventual clearance of infection.

A cardinal feature of T-cell immunity to infection is the activation, and subsequent transition of "naïve" T-cell precursors (i.e., cells that have not yet been exposed to their

https://doi.org/10.1016/B978-0-12-817964-2.00005-8
© 2020 Elsevier Inc.
All rights reserved.

pathogen and have no function) into effector T cells. Antigen recognition by the T-cell receptor induces a largely autonomous program of proliferation and differentiation that results in acquisition of lineage-specific T-cell functions required for control and elimination of an infection (Fig. 1).

Activated CD4$^+$ T cells can differentiate into several 'helper T-cell' lineages depending on the receipt of specific signals at the time of initial activation with each subset

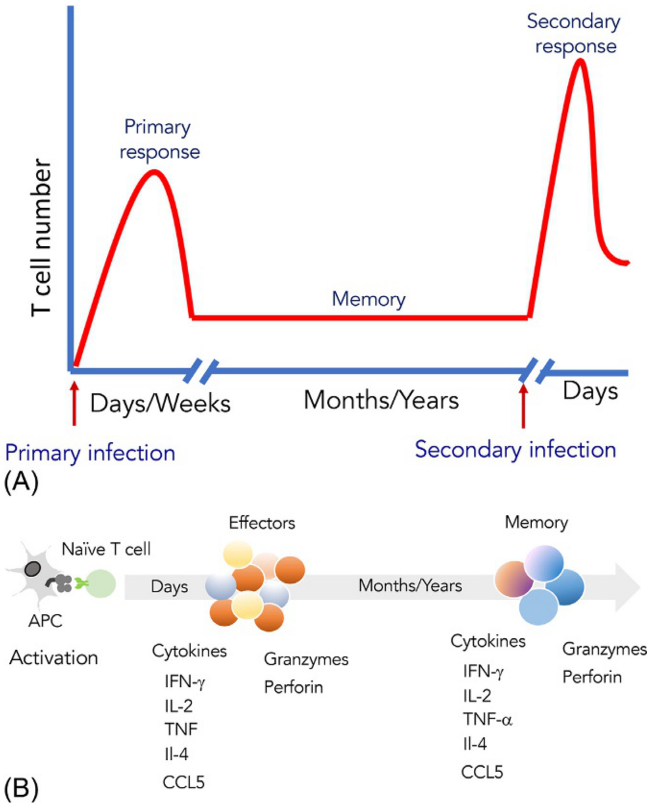

Secondary response
Primary response
T cell number
Memory
Days/Weeks
Months/Years
Days
Primary infection
Secondary infection
(A)
Effectors
Memory
Naïve T cell
Days
Months/Years
APC
Activation
Cytokines
IFN-γ
IL-2
TNF
Il-4
CCL5
Granzymes
Perforin
Cytokines
IFN-γ
IL-2
TNF-α
Il-4
CCL5
Granzymes
Perforin
(B)

Fig. 1 Primary and secondary Pathogen-specific T-cell responses. (A) Upon primary acute infection, T-cell activation initiates a program of proliferation that results in an increase in pathogen specific T-cell number. Once the infection is cleared, 90% of the effector T cells die off, leaving a small, but readily detectable population of memory T cells that can persist in the long term. Upon a secondary infection with the same pathogen, the increased cell numbers in a more expansive response. (B) The autonomous program of proliferation induced by T-cell activation also results in acquisition of T-cell lineage-specific effector functions. For CD4+ T cells this could include the production of cytokines such as IL-2, TNF, IFN-γ (T_H1), or IL-4 (T_H2). CD8+ T-cells express cytolytic molecules such as Granzyme (Gzm) A, Gzm B, and Perforin, as well as inflammatory cytokines such as IFN-γ, TNF, and the chemokine CCL5. Memory T cells retain this functional capacity upon reactivation after secondary infection without the need for further differentiation. *(Credit: Stephen J. Turner (Author).)*

exhibiting distinct transcriptional and functional characteristics. This in turn results in a tailored immune response against distinct pathogens. For example, naïve T_H-cell activation in the presence of the proinflammatory molecules, IFN-γ and IL-12, induces T_H1 differentiation, and these play a key role in controlling intracellular pathogens. Naive CD4+ T-cell activation in the presence of IL-4 generates a T_H2 response characterized by the production of IL-4, IL-5, and IL-13, which dictates the key role that these cells play in combating parasitic and helminth infections [1]. Importantly, the ability of extracellular signals to drive naive $CD4^+$ T_H-cell differentiation into distinct fates is dependent on specific induction of particular transcription factors (TFs) [2]. For example, T_H1 differentiation is dependent expression of the T-box TF T-BET (encoded by *Tbx21*; [3]). Conversely, IL-4 signals activate STAT6 resulting in upregulation of the TF GATA3 [4].

$CD8^+$ T cells express a range of key effector genes that equip them to mediate control of intracellular pathogens and cancer. This is via coordinated expression of variety effector molecules that include proinflammatory cytokines such as interferon (IFN)-γ, inflammatory chemokines such as CCL5), with concomitant expression of multiple effector molecules that mediate the cytotoxic activity, such as perforin and the granule enzymes (granzymes, GZM) A, B, and K (Fig. 1A) [5, 6]. Similar to T_H lineage determination, acquisition of signature killer cell function is mediated by specific transcription factors. For example, T-BET expression which is normally associated with $CD4^+$ T_H1 lineage commitment, in part, by promoting expression of IFN-γ [7], it is also rapidly upregulated in activated CTL and contributes to rapid acquisition of IFN-γ production and helps promote GZMB expression [8]. We recently showed that T-BET expression was required very early after activation for the initiation of appropriate lineage-specific $CD8^+$ T-cell differentiation and proliferation in response to infection [9].

Once an infection is cleared, most effector T cells die, leaving behind a small pool of long-lived cells that can recognize the same pathogen that triggered their initial activation (termed memory T cells) (Fig. 1B). Importantly, these memory T cells produce a broader array of immune molecules than naïve cells, and in larger quantities, and unlike naïve cells, can respond to infection without the need for further differentiation [10–12]. These features, combined with persistence at a higher frequency, enable memory T cells to respond more rapidly upon secondary infection, enabling earlier control and clearance of infection (Fig. 1B). These features of memory T cells provide the basis of T-cell–mediated immunity to reinfection [13, 14].

Functional heterogeneity within the memory T-cell pool

Phenotypic and functional heterogeneity within the memory T-cell pool result in distinct memory T cell roles in response to secondary challenge [15, 16]. Central memory T cells (T_{CM}) typically express the lymph node homing markers, CD62L (L-selectin) and CCR7 and exhibit extensive proliferative capacity upon reactivation [17]. More recently, another subset of memory T cells, also with proliferative capacity have been termed stem cell-like memory T cells (T_{SCM}) and are thought to reflect memory T cells that have significant capacity for self-renewal a greater recall capacity and can give rise to other effector and memory T-cell subsets [18–20].

In contrast, effector memory T cells (or T_{EM}) typically express tissue-specific homing markers such as CCR5, CXCR3, and integrins, and are capable of entering nonlymphoid tissues from the circulation in the steady state [17, 21, 22]. T_{EM} exhibit immediate effector function, such as cytotoxicity in the case of CTL [17] or cytokine production upon TCR ligation, without the need for further differentiation [23]. More recently, it has been demonstrated that memory T cells are established as tissue resident (termed resident memory T cells, T_{RM}) and able to persist in the long-term at the original site of infection. This provides a frontline defense against secondary infection at the primary pathogen entry site [24–26]. Together the different memory T-cell subsets provide layers of immune protection with T_{RM} providing the frontline defense against secondary infection. This can be further reinforced by recruitment of T_{EM} from the circulation, and expansion of T_{CM} in the draining lymph node that provides the extra numbers required to ensure complete control and elimination of an invading pathogen [27].

Our understanding of the molecular factors that shape cell-fate decisions and drive acquisition of T-cell effector function is limited, and questions remaining to be determined include how a T cell decides to be a memory vs an effector cell, and what are the molecular mechanisms that enable stable maintenance of rapid effector function within memory T cells in the long-term? In this chapter, we highlight those studies that address how epigenetic and transcriptional regulation controls acquisition and maintenance of lineage-specific effector function, and how it can impact results activated T-cell fate decisions.

Histone methylation and pattern, acquisition of function

Eukaryotic cells package and arrange their genetic information via formation of a DNA/histone protein complex called chromatin. The basic unit of chromatin is the nucleosome, an octameric protein complex consisting of two copies each of the core histone 2A (H2A), H2B, H3, and H4 proteins that DNA wraps around. Chromatin forms long fibers that can fold onto themselves to form chromosomes.

The composition of chromatin structure and biochemical modifications of histone proteins have emerged as important mechanisms for the regulation of gene transcription. Histone modifications in particular are known to regulate gene expression by either changing chromatin structure and/or by providing a platform that promotes binding of transcriptional regulators [28]. Histone proteins can be modified by a vast array of covalent modifications, particularly on the solvent N-terminal tail with the combination of posttranslational histone modifications (PTMs) and their genomic location act as a predictor of transcriptional activity [29, 30]. For example, acetylation of histone H3 at lysine 9 (H3K9Ac) that is associated with actively transcribed gene promoters [29]. Broadly speaking, histone acetylation is thought to promote an open chromatin structure by masking the overall positive charge of histones, thus reducing the intimacy of the interaction between acetylated nucleosomes and the (negatively charged) DNA [31]. Further, acetylation of consecutive nucleosomes results in charge repulsion, increasing the spacing between the nucleosomes, thereby increasing chromatin accessibility. In turn, increased accessibility potentiates transcription factor binding and the recruitment of the core transcription machinery.

Methylation of histone proteins is somewhat more complicated. For example, tri-methylation of histone 3 at lysine 4 (H3K4me3) is associated with almost all actively transcribed genes and has a strong correlation with histone acetylation and RNA polymerase II docking, indicative of a transcriptionally active genes, and genes that are "poised" for rapid activation [32–34]. However, how H3K4me3 deposition functions to regulate gene transcription is less clear than for histone acetylation, perhaps in part because it may function to maintain the transcriptional competency of some genes by excluding binding of repressive complexes, rather than by directly influencing transcriptional activation [35]. There is also evidence that H3K4me3 deposition influences rates of transcriptional elongation when it is deposited in broad domains [36]. In contrast to transcriptional activation, deposition of tri-methylation of H3K27 (H3K27me3) within promoter regions is linked to transcriptional repression [29, 37]. Another repressive mark, H3K9me3, is a particularly well-characterized modification implicated in heterochromatin formation through its interaction with heterochromatin protein 1 (HP1) [38, 39].

It is becoming increasingly evident that both the combination and extent of histone PTMs at specific genomic locations control gene transcriptional activity [29, 40, 41]. The combination of different active and repressive histone modifications at specific loci defines the transcriptional state of individual genes [42], thus defining cell fate. For example, studies in embryonic stem cells (ESCs) have demonstrated that transcriptional potential is maintained by co-enrichment of both active (H3K4me3) and repressive (H3K27me3) modifications at developmentally important gene loci (termed bivalent loci) [43–45]. Importantly, upon differentiation, the vast majority of bivalent loci in ESCs are permanently repressed by maintenance of H3K27me3 and loss of H3K4me3, ensuring that only cell lineage-specific genes are expressed [44]. Conversely, transcriptionally upregulated genes required for cell fate commitment lose H3K27me3 and retain H3K4me3 [43] (Fig. 2). These data suggest that bivalency is an epigenetic state from which a gene can be rapidly activated or repressed depending on the differentiation pathway initiated.

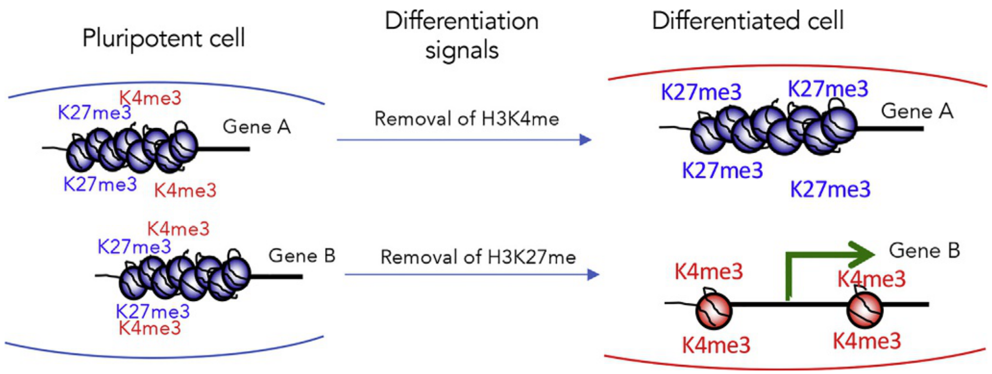


Fig. 2 The combination of histone PTMs at lineage-specific genes can influence cell fate decisions. (a) In pluripotent stem cells, many fate determining gene loci exhibit deposition of both the transcriptionally repressive, H3K27me3 and permissive H3K4me3 histone PTMs. Upon receipt of specific differentiation signals, genes required for the commitment toward a specific cell fate lose H3K27me3 and retain H3K4me3, resulting in rapid transcriptional upregulation. Genes required for a distinct cell fate lose H3K4me3 and retain H3K27me3 essentially shutting down that gene. This ensures only appropriate cell specific gene transcription of cell fate determining genes. *(Credit: Stephen J. Turner (Author).)*

DNA methylation and its role in regulating gene transcription

The addition of methyl groups to cytosine residues associated with CpG dinucleotides is a major epigenetic mechanism enabling heritable gene silencing [46]. It has long been appreciated that DNA methylation is an important mechanism for regulating mammalian embryonic development and subsequent cellular differentiation [47]. The addition of and removal of DNA methylation is a dynamic process that involves DNA methyltransferases such as DNMT3A and DNMT3B [48]. Conversely, TET enzymes that facilitated the oxidation of 5-methyl cytosine (5mC) to 5-hydroxmethylcytosine (5hmC) [49]. This conversion from 5mC to 5hmC then facilitates full demethylation at genomic regions and is associated with transcriptional activation. Within the mammalian genome, CpG dinucleotides are enriched within specific genomic regions, termed CpG islands, that are often located within mammalian gene promoters. The subsequent methylation/demethylation of such regions serve to regulate gene transcription [50].

Making sense of the "junk DNA": Noncoding regulatory elements work via chromatin folding

As little as 2% of the human genome encodes for proteins with the rest of the DNA once considered largely redundant and nonfunctional and was termed "junk DNA." What we now know is that much of this noncoding DNA contains important regulatory elements, termed "transcriptional enhancers" (TEs) [51]. TEs are short sequences of DNA that regulate temporal and spatial patterns of gene transcription, and as such, TEs are key determinants of cellular differentiation. These regulatory DNA elements can be positioned upstream, downstream (often many hundreds of kilobases distant) or even within target genes that they regulate. TEs convey their effects on gene transcription by binding transcription factors, which they bring into close proximity to their cognate gene promoter by chromatin folding. Moreover, there is emerging evidence that gene loci can be regulated by more than one TE [51]. The chromatin structure and associated biochemical chromatin modifications at TEs are highly dynamic with changes at these elements a major regulatory mechanism controlling cellular differentiation and fate commitment [52]. It is now becoming apparent that the importance of chromatin remodeling in regulating gene expression during cellular differentiation is in large part via promotion or disruption of TE:target gene contacts.

Epigenetic regulation in the acquisition of lineage-specific T-cell function

Prior to activation, naïve T cells are quiescent and lack immediate effector function. As little as 2 h of initial stimulation is sufficient to initiate a program of proliferation and differentiation [53]. For both CD4$^+$ and CD8$^+$ T cells, the acquisition of effector function is clearly linked with cellular proliferation [54–58]. In addition, a strict hierarchy of cytokine production is also observed for virus-specific CD8$^+$ T cells that likely reflects sequential acquisition of multiple effector functions due to progressive differentiation [11, 59]. Given that distinct differentiation states within both CD4$^+$ and CD8$^+$ T cells, are associated within distinct phenotypic and functional states, how changes in the epigenetic landscapes impact acquisition and maintenance of lineage- and differentiation-specific effector function has been of great interest.

There has been extensive analysis of the specific epigenetic mechanisms that underpin CD4$^+$ effector T-cell lineage commitment from a naïve state into different effector subsets (reviewed by Ansel et al. [4] and Wilson [60]). Initial studies focused on epigenetic changes that result in expression of IFN-γ (encoded by *Ifng*) or IL-4/IL-5/IL-13 (encoded by *Il4, Il5,* and *Il13*) loci, within T$_H$1 vs T$_H$2 CD4$^+$ T cells, respectively [61]. Differentiation of naïve CD4$^+$ T cells into either lineage results in simultaneous chromatin remodeling at both the *Ifng* and *Il4/Il5/Il13* loci with differing outcomes. For example, T$_H$2 differentiation results in removal of DNA methylation, increase chromatin accessibility, removal of H3K27me3 and deposition of H3K4me3 and H3K9/27 acetylation at key regulatory elements within the *Il4/Il5/Il13* locus [62, 63]. Conversely, the *Ifng* locus also undergoes extensive remodeling with an increase in CpG methylation and H3K27me3 deposition resulting in the repression of gene transcription [64]. Importantly, these alternate outcomes to chromatin remodeling within a T$_H$2 cells are both dependent on the action of GATA3 and STAT6 transcription factors [65, 66]. This highlights how TFs networks can seemingly have opposing functions within the same cell, but nevertheless, combine to reach the same outcome, in this case commit cells to the T_H2 lineage.

Differentiation of naive $CD4^+$ T cells into T_H1 cells results in establishment of a permissive chromatin landscape at the *Ifng* locus. This includes increased H3K4me3 and histone acetylation upon activation, increased chromatin accessibility and removal of repressive histone PTMs (H3K27me3) [61, 63, 67]. Acquisition of T_H1 effector function is dependent on the upregulation of T-BET [7, 68], as well as STAT4 (via IL-12 signaling) [69] and STAT5 (via IL-2 signaling) [70]. As before, T-BET expression also serves to bind to and shutdown both GATA3 and IL-4/5/13 expression during T_H1 differentiation by pairing with other TFs, such as RUNX3, and helping establish a repressive chromatin landscape [3, 71]. Interestingly, T-BET can also directly interact with GATA3 to inhibit its role in T_H2 cytokine activation further reinforcing CD4+ T_H1 fate commitment [71].

Epigenetic regulation in the acquisition of CD8 T-cell function

As with acquisition of $CD4^+$ lineage-specific function, dynamic changes in DNA methylation status and specific histone modifications underpin observed phenotypic and functional changes during $CD8^+$ T-cell differentiation [72]. Early studies demonstrated that naïve $CD8^+$ T cells exhibit a transcriptionally repressive epigenetic signature at the promoter of genes such as the *Ifng*, *Gzmb*, *Gzma*, and *Pfp* and other effector gene loci that mediate the signature cytotoxic CTL function [59, 73, 74].

Several genome wide studies have since examined changes in DNA methylation, chromatin accessibility, and histone PTMs laying out detailed insights into the dynamics of chromatin remodeling during pathogen-specific $CD8^+$ T cell [75–82]. As expected, acquisition of lineage-specific effector function upon CTL activation was associated with the establishment of a permissive chromatin landscape as described above (Fig. 3). Importantly, the establishment of a permissive chromatin landscape at CTL effector gene loci (e.g., *Ifng* and *Fasl*) was maintained into the memory state [75–82]. Moreover, binding of RNA polymerase II could also be demonstrated in memory effector gene loci with this transcriptional poising enabling rapid gene expression upon reactivation, a characteristic of memory CTL [59, 74, 83] (Fig. 3). As such, the maintenance of transcriptionally poised effector gene loci in memory CTL is a molecular mechanism underpinning a cardinal feature of T-cell memory, namely, rapid effector function upon reactivation [84].

While CD8+ T-cell effector gene loci exhibited a classic switch from a transcriptionally repressive ($H3K27me3^+H3K4me3^-$) to transcriptionally permissive ($H3K27me3^-$ $H3K4me3^+$) chromatin landscape, closer inspection identified a more complex dynamic. In particular, transcriptional activation of a broad array of genes upon T-cell activation was linked to the removal of H3K27me3 at gene promoters already pre-marked with H3K4me3 in the naïve state (Fig. 4B) [76, 77]. Antigen-dependent T-cell activation resulted in removal of H3K27me3 at many gene promoters prior to the first cell division.

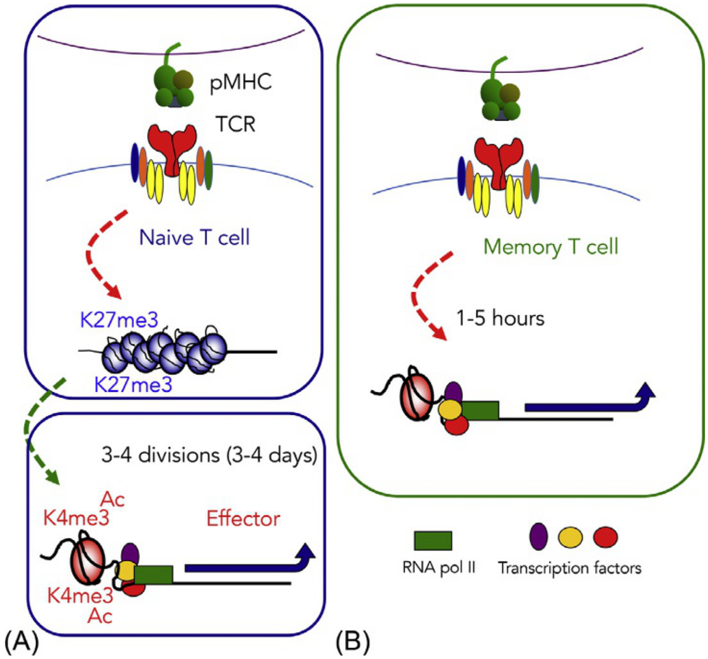


Fig. 3 Acquisition and maintenance of CD8^{+} T-cell effector function is associated with chromatin remodeling and establishment of a transcriptionally poised chromatin state (A) The effector genes within naïve CD8+ T cells exhibit a transcriptionally repressive chromatin landscape. Upon T-cell activation, effector gene promoters undergo an increase in chromatin accessibility, loss of H3K27me3, and gain of H3K4me3/histone acetylation. This process takes 1–3 days during the proliferative response. Transcriptional machinery such as TFs, RNA polymerases and elongation factors are recruited to the transcriptional start site to initiate gene transcription. (B) Effector gene loci within pathogen-specific memory T cells maintain a transcriptionally permissive chromatin landscape around gene promoters. Moreover, many of these gene loci have components of the transcriptional machinery docked onto the chromatin. Upon receipt of T-cell activation signals, gene transcription occurs rapidly, without the need for further differentiation. *(Credit: Stephen J. Turner (Author).)*

Unlike CD8^{+} T-cell effector gene loci, this results in the establishment of a permissive transcriptional chromatin landscape without the need for de novo H3K4me3 deposition (Fig. 4B and C). Interestingly, it was found that genes regulated in this manner played key roles in processes required for sustaining a rapid proliferative cellular response upon T-cell activation [76]. This configuration is reminiscent of the "bivalency" describe in embryonic stem cells described above (Fig. 2). Moreover, a subset of these gene loci exhibited co-deposition of H3K4me3 and H3K27me3 on the same histone tail [76], so were truly bivalent (Fig. 4C). In this case, this configuration marked specific TFs, such as *Tbx21* and *Irf4*, both key for initiating and sustaining optimal CD8+ T-cell responses [8, 85]. A similar observation has been made in CD4+ T cells where co-localization of both the permissive and repressive histone PTM is evident, particularly at the *Tbx21*

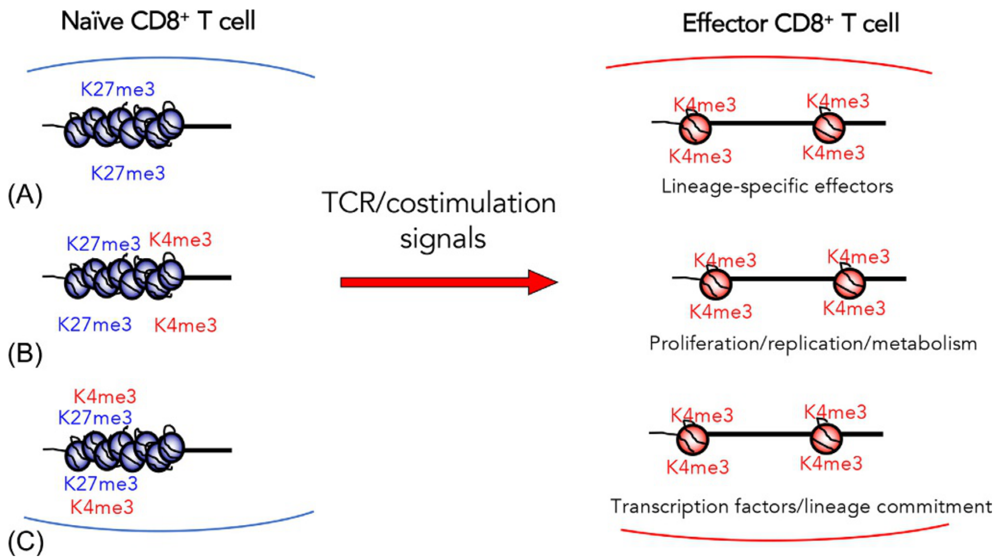

Naïve CD8+ T cell
Effector CD8+ T cell
K27me3
K27me3
(A)
K27me3 K4me3
K27me3 K4me3
(B)
K4me3
K27me3
K27me3
K4me3
(C)
TCR/costimulation signals
K4me3 K4me3
K4me3 K4me3
Lineage-specific effectors
K4me3 K4me3
K4me3 K4me3
Proliferation/replication/metabolism
K4me3 K4me3
K4me3 K4me3
Transcription factors/lineage commitment

Fig. 4 Distinct histone methylation dynamics is associated with regulating genes in different cellular functions (A) Within naïve T cells, $CD8^+$ T-cell effector gene loci exhibit a transcriptionally repressive chromatin landscape that transition to a transcriptionally permissive chromatin structure upon activation. (B) Within naïve CD8+ T cells, a second class of gene loci exhibit enrichment of both the repressive and permissive histone PTMs. Upon T-cell activation, there is rapid loss of H3K27me3 and retention of H3K4me3. This enables rapid transcriptional activation of genes involved in broad cellular processes such as cell proliferation, DNA replication, and cellular metabolism. These processes likely help sustain T-cell proliferative responses. (C) A third class of gene loci exhibit colocalization of both H3K27me3 and H3K4me3 PTMs on the same histone tail. These genes are enriched for CD8+ T-cell transcription factors such as *Tbx21* and *Irf4*. H3K27me3 is removed and H3K4me3 retained at these gene loci upon T-cell activation. This is associated with rapid transcriptional activation prior to first-cell division. *(Credit: Stephen J. Turner (Author).)*

locus [63]. Exactly how these distinct histone patterns (or codes) are recognized to enact H3K27me3 removal is not completely clear. Nevertheless, it appears that H3K27me3 deposition is able to override any permissive histone code and acts as a molecular hand brake within naïve T cells to restrain T-cell gene transcription. T-cell activation results in the rapid removal of H3K27me3 ensuring the rapid transcriptional upregulation of genes that able to ensure fate commitment, but also support the autonomous program of cellular differentiation into the effector/memory state.

Epigenetic mapping of the differentiation pathway that leads to T-cell memory

While the establishment and maintenance of permissive chromatin structures at effector gene loci within memory T cells ensures rapid effector function, what is still unclear is whether modulation to epigenetic modifications also play a role in determining the fate of recently activated T cells. Recent studies have examined how the epigenetic landscape changes as naïve T cells differentiate into different effector and memory subsets. Genome-wide profiling of histone PTMs of human memory T-cell subsets indicated that there may in fact be a gradual pathway of cellular differentiation, with the naïve and effector states representing the start and end of this process, respectively [86]. For example, expression of TCF-1 (encoded by *Tcf7*) is key for maintenance of the naïve T-cell state and is associated with a permissive epigenetic chromatin landscape at the *Tcf7* gene locus within naïve T cells [81, 86]. Comparison of the chromatin landscape within different CD8$^+$ T-cell memory and effector subsets showed that there were small but progressive changes to the epigenetic landscape during the transition of naïve CD8$^+$ T cells to the memory/effector state. In fact, graded epigenetic silencing of the *Tcf7* gene locus correlated with progressive TCF-1 downregulation suggesting a linear differentiation pathway with a naïve → T_{SCM} → T_{CM} → T_{EM} → effector transition [81, 86]. This same process was observed in other signature "naïve" gene loci such as *Lef1* and *Foxo1* [81, 86]. Similarly, gradual remodeling of the chromatin landscape from a repressive to permissive state at CD8$^+$ effector gene loci was also evident during mouse CD8$^+$ pathogen-specific T-cell differentiation [76, 81, 86]. These studies suggest that T-cell activation initiates molecular programs that simultaneously remodel the chromatin landscape at functionally distinct gene loci into either transcriptionally permissive (effector genes) or repressive states (cell fate genes). The specific targeting of distinct genes with distinct epigenetic mechanisms is most likely controlled by transcription factor targeting of specific genes and their associated regulatory elements.

The role of CD8$^+$ T-cell-specific transcription factors in chromatin remodeling and acquisition of function

Genome wide studies mapping changes in chromatin accessibility, or changes in histone PTMs, have demonstrated that the specific genomic regions undergoing remodeling are associated with the enrichment of particular transcription factor binding sites [76–79, 81, 87]. A number of TFs, such as IRF4 [85], T-BET [8], RUNX3 [8, 87], ZEB2 [88, 89], PRDM-1 [90, 91], STAT5 [92], and GATA3 [93, 94], have all been shown to be important for initiating, promoting, and stabilizing signature effector CD8$^+$ T-cell function. Nevertheless, the precise mechanism by which such TFs contribute to the remodeling of the epigenetic landscape within effector CD8$^+$ T cells is not as well studied.

It has been previously demonstrated that T-BET can physically interact with and recruit H3K27 demethylases to the *Ifng* regulatory elements in CD4$^+$ T_H1 cells, but interestingly, the function of the demethylases is not to demethylate the histones, but rather to serve as a structural scaffold [95]. The removal of H3K27me3 itself may actually be done by evicting the nucleosomes baring the modification from the locus through the T-BET-dependent recruitment of SWI/SNF remodeling complexes [96]. We have recently demonstrated that T-BET-deficient effector cells failed to remodel *Ifng* associated TEs from a poised to active state, which is consistent with the finding that T-BET also recruits H3K4 methyltransferase activity to the locus [9]. Moreover, chromatin remodeling and activation of the *Ifng* locus within CD4$^+$ T cells has also been shown to involve T-BET–dependent recruitment of the CCCTC-binding factor (CTCF) to regulatory elements [97]. In an earlier report, it was demonstrated that early NFAT and AP-1 upregulation contributed to establishing a poised chromatin signature at specific TEs associated with regulation of T-cell differentiation [83]. Hence, TFs can recruit specific chromatin remodeling proteins to specific genomic regions, to either help initiate initial changes to chromatin structure, or to help stabilize them.

In terms of driving memory vs effector fate decisions, recent work has focused on the early changes in chromatin accessibility, particularly at regulatory elements, and the associated TFBSs enriched within these restructured chromatin regions [77, 79, 81, 87]. With the use of novel algorithms to identify and rank potential TFBSs, it is now possible to make predictions regarding the importance of key TFs involved in determining either CD8$^+$ T-cell effector or memory fate decisions [78, 79, 81]. Comparison of distinct chromatin accessibility between CD8$^+$ T-cell effector and memory precursors generated after bacterial infection showed enrichment for already known regulators of effector CD8+ T-cell differentiation such as BATF, AP-1, and T-BET [81]. Importantly, CD8$^+$ memory precursors showed enrichment for TCF-1, LEF-1, and E2A at accessible chromatin regions compared to CD8$^+$ T-cell effector cells [81]. These same TFs are also known to be important for the maintenance of naïve CD8$^+$ T cells [98]. This supports the earlier notion that naïve and memory CD8$^+$ T cells share a similar epigenetic landscape [83], and hence rely on a similar suite of TFs for the maintenance of their quiescent state [99, 100]. In the same study, the results show a previously underappreciated role for the TFs YY-1 and NR3C1 (glucocorticoid receptor) in regulating either CD8$^+$ T-cell effector or memory fate decisions, respectively [81]. Upon corticosteroid binding, NR3C1 translocates to the nucleus to act as a transcriptional transactivator [101]. Treatment of mice with glucocorticoids during bacterial infection promoted the formation of memory CD8$^+$ T cells [81]. Hence, the use of novel computational approaches to interrogate the epigenetic landscape of pathogen-specific CD8$^+$ T cells has the potential to not only identify novel TFs involved in CD8$^+$ T-cell fate decisions, but also the associated signaling pathways upstream of these TFs that determine effector vs memory cell fate decisions.

A similar approach examining response to viral infection also compared regions of chromatin that became accessible soon after T-cell activation. In this study, they identified that many of the regions that became accessible upon activation were maintained into memory [87]. Analysis of TFBS-binding enrichment demonstrated that many of these regions possessed RUNX family member-binding motifs. RUNX3 expression, a TF that had been demonstrated previously to be a key TF for ensuring fidelity of CD8$^+$ T-cell function upon activation [8, 102], was required for the dynamic changes in chromatin accessibility observed soon after naïve T-cell activation [87]. Further, RUNX3 binding was already evident within naïve CD8$^+$ T cells. Together these data suggest that RUNX3 likely acts as pioneering factor whose role is to remodel chromatin at CD8$^+$ T-cell-lineage-specific gene loci, making these regulatory elements accessible for other CD8$^+$ T-cell TFs such as BATF, IRF4, and BLIMP1 [87].

Interestingly, RUNX3 deficiency resulted in a loss of memory CD8$^+$ T-cell formation and promotion of effector T-cell subsets [87]. Both T-BET and its downstream target ZEB2 were repressed after RUNX3 overexpression. Hence, RUNX3 not only serves to set up an appropriate chromatin landscape enabling effective CD8+ T-cell differentiation, but it also plays a key role in promoting memory formation via limiting effector differentiation [87]. More studies that focus on the precise molecular mechanisms of TF activity will undoubtedly provide key information about their role in chromatin remodeling leading to CD8+ T-cell memory formation.

Active regulation of chromatin state is a key factor in CD8$^+$ T-cell effector vs memory fate decisions

The fact that RUNX3 is involved in multiple mechanisms ensuring optimal CD8$^+$ T-cell differentiation and memory formation is reminiscent of T-BET and GATA3 playing dual roles that when combined ensure appropriate fate decisions after activation. A question that remains is exactly when a recently activated CD8$^+$ T cell is directed to become either and effector or memory cell? Use of single-cell approaches have demonstrated that there is molecular heterogeneity observed after the first-cell division. In fact, there was separation of cell populations that had differential gene signatures indicative of either the effector or memory fate [103, 104]. For example, a proportion of cells that had undergone a single cell division, and exhibited higher levels of genes involved in CD8$^+$ T-cell effector function, metabolic reprogramming and cell division, were shown to be enriched in cells that became effector cells [103]. Conversely, the alternate population was enriched for genes associated with memory formation and not suprising contained memory T cell precursors.

Analysis of single-cell RNA-seq profiles of responding CD8$^+$ T cells also identified novel potential regulators of T-cell differentiation and effector vs memory commitment [104]. One of these regulators was Enhancer of Estes homolog-2 (EZH2), the catalytic component of the Polycomb Repressor Complex 2 (PRC2) which mediates the methylation of H3K27 [105, 106]. EZH2, and other components of the PRC2 complex, were found to be more highly expressed within CD8$^+$ T cells that had undergone one cell division and were destined to become effector cells vs memory cells [104]. ChIP-seq analysis demonstrated that pro-memory and pro-survival gene loci within effector CD8$^+$ T cells exhibited higher levels of H3K27me3 compared to memory [104, 107]. This suggested that PRC2 likely acts to repress genes that promote memory formation and retain proliferative potential.

This was supported by analysis of EZH2-deficient CD8$^+$ T-cell responses to either bacterial or viral infection [104, 107]. In both cases, EZH2-deficient CD8$^+$ T cells failed to generate substantial effector T-cell populations. Hence, PRC2 dependent addition of H3K27me3 to pro-memory gene loci represents a key decision point that results in a switch from promoting memory formation to promoting effector CTL differentiation.

A more recent study focused on another histone methyltransferase, SUV39H1, which mediates the methylation of H3K9, a transcriptionally repressive histone PTM on the formation of CD8$^+$ T-cell effector and memory differentiation [108]. SUV39H1 had already been described in ensuring lineage commitment of recently activated T_H2 CD4$^+$ T cells [109]. SUV39H1 deficient T_H2 CD4$^+$ T cells were able to express T_H1 type effector genes upon re-culture in T_H1-stimulating conditions [109]. Hence, SUV39H1-dependent H3K9 methylation was required for epigenetic silencing of T_H1 cell fate. In the case of CD8$^+$ T-cell responses to infection, CD8$^+$ T-cell–specific SUV39H1 deficiency resulted in a diminished protection from bacterial challenge [108]. Reminiscent of findings examining the role of PRC2 complex in antiviral CD8$^+$ T-cell responses, it was found that SUV39H1 deficient CD8$^+$ T cells could not epigenetically silence pro-stem cell and pro-memory genes such as *Il7ra*, *Sell*, and *Tcf7* (Fig. 5). In this case, SUV39H1 CD8$^+$ T cells are not able to fully transition from a naïve/memory state and fully engage effector program.

While the effector-memory fate decision requires deposition of the repressive H3K27me3 and H3K9me3 modifications, removal of these same marks occurs simultaneously at other parts of the genome. We have demonstrated that removal of H3K27me3 represents a major change in chromatin landscape between naïve and memory/effector T cells [76]. The question is just how then are opposing demethylases and methyltransferases targeted to distinct genomic targets during the process of T-cell activation and differentiation? The removal of H3K27me3 is specifically catalyzed by KDM6A and KDM6B demethylases [110] and it has been associated with acquisition of T-cell effector function. Whether this also contributes to the memory vs effector fate commitment is less clear. In mature CD4$^+$ T cells, KDM6A activity was required for the rapid expression of several key transcription factors, such as TBET and STAT family members [111]. *Kdm6b*-deficient CD4$^+$ T cells demonstrate dysregulated and inappropriate fate specification under T_H skewing conditions with promotion of T_H2/T_H17 lineages at the expense of T_H1 and FOXP3$^+$ T regulatory cells [76]. Whether KDM6A/B removal is there to help prime the chromatin landscape for subsequent molecular checkpoints, or whether there is an active targeting of gene loci that promote effector differentiation remains to be determined.

In line with earlier observations for PRC2- and SUV39H1-dependent transcriptional repression, upregulation of the DNA methyltransferase, DMNT3A, upon virus-specific CD8$^+$ T-cell activation was associated with increased CpG methylation at pro-memory/naïve gene loci and subsequent promotion of effector T-cell differentiation [112]. This

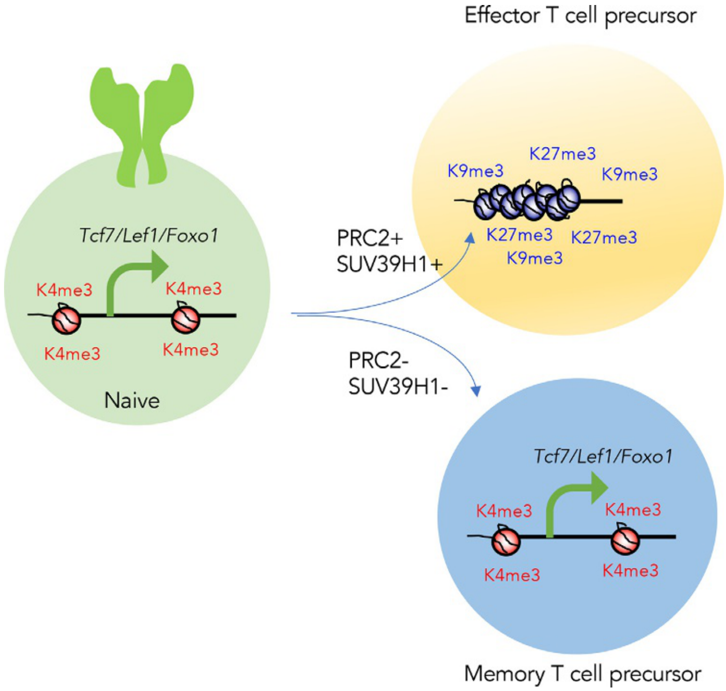

Effector T cell precursor
K27me3
K9me3
K9me3
K27me3
K27me3
K9me3
PRC2+
SUV39H1+
Tcf7/Lef1/Foxo1
K4me3
K4me3
K4me3
K4me3
Naive
PRC2-
SUV39H1-
Tcf7/Lef1/Foxo1
K4me3
K4me3
K4me3
K4me3
Memory T cell precursor

Fig. 5 The role of chromatin-modifying enzymes in directing effector vs memory T-cell fate decisions (a) Genes associated with maintaining pluripotency within naïve T cells such as *Tcf7*, *Lef1*, and *Foxo1* are actively transcribed. Upon T-cell activation and cell division, the histone methyltransferases *Ezh2* (a member of the polycomb repressive complex 2) or SUV39H1 can be upregulated. If a cell expresses these histone methyltransferases, they can target and shut down *Tcf7, Lef1*, and *Foxo1* transcription via deposition H3K27me3 and/or H3K9me3. Shut down of these pluripotency genes promotes effector T-cell differentiation. Conversely, if a cell does not express these HTMs, *Tcf7*, *Lef1*, and *Foxo1* transcription is maintained promoting memory T-cell formation. *(Credit: Stephen J. Turner (Author).)*

again supports the notion that imposition of repressive epigenetic mechanisms serves to inactivate the naïve T-cell program, enabling activated T cells to transition into the effector/memory differentiation state. Specific deletion of *Dnmta* within activated CD8$^+$ T cells resulted in an enrichment of central memory T cells, indicative of a less differentiated state [83] and an inability to epigenetically shut down naïve/memory gene programs.

Another point to consider is that deposition of H3K27me3 and H3K9me3 at gene loci controlling self-renewal/pro-memory formation in CD8$^+$ T cells would presumably establish an epigenetic barrier representing a point of no return for differentiating T cells. That is to say, CD8$^+$ T cells that transition from the naïve/memory state to an effector state would now be fully committed to what would effectively be a terminal fate. Intriguingly, recent comparison of effector CD8$^+$ T cells showed that apparent epigenetic silencing of naïve/memory gene loci, could be reversed enabling reexpression of memory genes [112]. Transfer of CD62L^{lo} effector cells resulted in the generation of both CD62L^{hi} and CD62L^{lo} populations. Analysis of the *Sell* (encoding CD62L) showed that the reexpression of CD62L by a subset of the memory CD8$^+$ T cells coincided with demethylation at the *Sell* locus [112]. These data suggest that memory formation might indeed occur via dedifferentiation of effector CD8$^+$ T cells [82, 112]. What was not clear from these studies is whether there was also dedifferentiation at other gene loci (such as *Tcf7* or *Lef1*). Hence, whether this is true restoration of the pluripotent/memory program is unclear. Further, other data suggest that the ability to reprogram virus-specific CTL is really quite dependent on the extent of differentiation [79, 113, 114]. Memory CTL that have undergone extended differentiation are not readily able to be reprogrammed, and hence, their cellular fate is fixed [113, 114]. Finally, given the analysis was done on bulk populations, there is also the possibility that the emergence of CD62L^{hi} memory CTL from the CD62^{lo} effector population might represent enrichment of precursors already present, but not readily detected. The use of single-cell epigenetic approaches would be able to delineate the extent of cellular heterogeneity within these populations and give some insight into whether this is the case.

Conclusion

There has long been an appreciation of how alterations to chromatin structure are associated with naïve T-cell activation and the acquisition and maintenance of lineage-specific function. Recent advances in genomic approaches and computational biology have provided greater insights into both the dynamics of epigenetic modulation and potential role of specific transcription factors in driving changes in the epigenetic landscape. The establishment of transcriptionally permissive chromatin structures at T-cell lineage effector genes enables acquisition of lineage-specific function. We now know this reprogramming of the epigenetic landscape is established early after infection, and further stabilized as the effector T-cell responses develop. Areas that still remain controversial are whether there is in fact dedifferentiation of effector cells into memory; or whether these landscapes once stabilized are in fact immovable. While not covered here, how genome wide changes in the three-dimensional architecture of chromatin enables acquisition and maintenance of T-cell-lineage-specific function is an area that is relatively understudied. Research into this, and other concepts such as how chromatin modifying enzymes are targeted to specific genomic regions, and just how uniform such changes are across a heterologous cellular population represent major areas of focus for the future.

References

[1] Zhu J, Yamane H, Paul WE. Differentiation of effector CD4 T cell populations (*). Annu Rev Immunol 2010;28:445–89.
[2] Kanno Y, et al. Transcriptional and epigenetic control of T helper cell specification: molecular mechanisms underlying commitment and plasticity. Annu Rev Immunol 2012;30:707–31.
[3] Djuretic IM, et al. Transcription factors T-bet and Runx3 cooperate to activate Ifng and silence Il4 in T helper type 1 cells. Nat Immunol 2007;8(2):145–53.
[4] Ansel KM, Lee DU, Rao A. An epigenetic view of helper T cell differentiation. Nat Immunol 2003; 4(7):616–23.
[5] Jenkins MR, et al. Heterogeneity of effector phenotype for acute phase and memory influenza A virus-specific CTL. J Immunol 2007;179(1):64–70.
[6] Peixoto A, et al. CD8 single-cell gene coexpression reveals three different effector types present at distinct phases of the immune response. J Exp Med 2007;204(5):1193–205.
[7] Szabo SJ, et al. A novel transcription factor, T-bet, directs Th1 lineage commitment. Cell 2000; 100(6):655–69.
[8] Cruz-Guilloty F, et al. Runx3 and T-box proteins cooperate to establish the transcriptional program of effector CTLs. J Exp Med 2009;206(1):51–9.
[9] Prier JE, et al. Early T-BET expression ensures an appropriate CD8(+) lineage-specific transcriptional landscape after influenza A virus infection. J Immunol 2019;203(4):1044–54.
[10] Kaech SM, et al. Molecular and functional profiling of memory CD8 T cell differentiation. Cell 2002;111(6):837–51.
[11] La Gruta NL, Turner SJ, Doherty PC. Hierarchies in cytokine expression profiles for acute and resolving influenza virus-specific CD8+ T cell responses: correlation of cytokine profile and TCR avidity. J Immunol 2004;172(9):5553–60.
[12] Veiga-Fernandes H, et al. Response of naive and memory CD8+ T cells to antigen stimulation in vivo. Nat Immunol 2000;1(1):47–53.
[13] Barber DL, Wherry EJ, Ahmed R. Cutting edge: rapid in vivo killing by memory CD8 T cells. J Immunol 2003;171(1):27–31.
[14] Kersh EN, et al. TCR signal transduction in antigen-specific memory CD8 T cells. J Immunol 2003;170(11):5455–63.
[15] Sallusto F, Geginat J, Lanzavecchia A. Central memory and effector memory T cell subsets: function, generation, and maintenance. Annu Rev Immunol 2004;22:745–63.
[16] Sallusto F, et al. Two subsets of memory T lymphocytes with distinct homing potentials and effector functions. Nature 1999;401(6754):708–12.
[17] Masopust D, et al. Preferential localization of effector memory cells in nonlymphoid tissue. Science 2001;291(5512):2413–7.
[18] Gattinoni L, et al. A human memory T cell subset with stem cell-like properties. Nat Med 2011; 17(10):1290–7.
[19] Gattinoni L, et al. Wnt signaling arrests effector T cell differentiation and generates CD8+ memory stem cells. Nat Med 2009;15(7):808–13.
[20] Lugli E, et al. Superior T memory stem cell persistence supports long-lived T cell memory. J Clin Invest 2013;123(2):594–9.
[21] Kohlmeier JE, et al. Inflammatory chemokine receptors regulate CD8(+) T cell contraction and memory generation following infection. J Exp Med 2011;208(8):1621–34.
[22] Wakim LM, et al. Cutting edge: local recall responses by memory T cells newly recruited to peripheral nonlymphoid tissues. J Immunol 2008;181(9):5837–41.
[23] Marshall DR, et al. Measuring the diaspora for virus-specific CD8+ T cells. Proc Natl Acad Sci U S A 2001;98(11):6313–8.
[24] Gebhardt T, et al. Memory T cells in nonlymphoid tissue that provide enhanced local immunity during infection with herpes simplex virus. Nat Immunol 2009;10(5):524–30.
[25] Gebhardt T, et al. Different patterns of peripheral migration by memory CD4+ and CD8+ T cells. Nature 2011;477(7363):216–9.

[26] Mackay LK, et al. Long-lived epithelial immunity by tissue-resident memory T (TRM) cells in the absence of persisting local antigen presentation. Proc Natl Acad Sci U S A 2012;109(18):7037–42.
[27] Wherry EJ, et al. Lineage relationship and protective immunity of memory CD8 T cell subsets. Nat Immunol 2003;4(3):225–34.
[28] Kouzarides T. Chromatin modifications and their function. Cell 2007;128(4):693–705.
[29] Wang Z, et al. Combinatorial patterns of histone acetylations and methylations in the human genome. Nat Genet 2008;40(7):897–903.
[30] Zhang Y, Reinberg D. Transcription regulation by histone methylation: interplay between different covalent modifications of the core histone tails. Genes Dev 2001;15(18):2343–60.
[31] Aoyagi S, et al. Nucleosome remodeling by the human SWI/SNF complex requires transient global disruption of histone-DNA interactions. Mol Cell Biol 2002;22(11):3653–62.
[32] Bernstein BE, et al. Genomic maps and comparative analysis of histone modifications in human and mouse. Cell 2005;120(2):169–81.
[33] Heintzman ND, et al. Distinct and predictive chromatin signatures of transcriptional promoters and enhancers in the human genome. Nat Genet 2007;39(3):311–8.
[34] Santos-Rosa H, et al. Active genes are tri-methylated at K4 of histone H3. Nature 2002; 419(6905):407–11.
[35] Nishioka K, et al. Set9, a novel histone H3 methyltransferase that facilitates transcription by precluding histone tail modifications required for heterochromatin formation. Genes Dev 2002;16(4):479–89.
[36] Chen K, et al. Broad H3K4me3 is associated with increased transcription elongation and enhancer activity at tumor-suppressor genes. Nat Genet 2015;47(10):1149–57.
[37] Barski A, et al. High-resolution profiling of histone methylations in the human genome. Cell 2007; 129(4):823–37.
[38] Lachner M, et al. Methylation of histone H3 lysine 9 creates a binding site for HP1 proteins. Nature 2001;410(6824):116–20.
[39] Jacobs SA, Khorasanizadeh S. Structure of HP1 chromodomain bound to a lysine 9-methylated histone H3 tail. Science 2002;295(5562):2080–3.
[40] Jenuwein T, Allis CD. Translating the histone code. Science 2001;293(5532):1074–80.
[41] Berger SL. The complex language of chromatin regulation during transcription. Nature 2007; 447(7143):407–12.
[42] Strahl BD, Allis CD. The language of covalent histone modifications. Nature 2000;403(6765):41–5.
[43] Bernstein BE, et al. A bivalent chromatin structure marks key developmental genes in embryonic stem cells. Cell 2006;125(2):315–26.
[44] Hawkins RD, et al. Distinct epigenomic landscapes of pluripotent and lineage-committed human cells. Cell Stem Cell 2010;6(5):479–91.
[45] Orford K, et al. Differential H3K4 methylation identifies developmentally poised hematopoietic genes. Dev Cell 2008;14(5):798–809.
[46] Colot V, Rossignol JL. Eukaryotic DNA methylation as an evolutionary device. Bioessays 1999; 21(5):402–11.
[47] Sanford JP, et al. Differences in DNA methylation during oogenesis and spermatogenesis and their persistence during early embryogenesis in the mouse. Genes Dev 1987;1(10):1039–46.
[48] Okano M, et al. DNA methyltransferases Dnmt3a and Dnmt3b are essential for de novo methylation and mammalian development. Cell 1999;99(3):247–57.
[49] Tahiliani M, et al. Conversion of 5-methylcytosine to 5-hydroxymethylcytosine in mammalian DNA by MLL partner TET1. Science 2009;324(5929):930–5.
[50] Smith ZD, Meissner A. DNA methylation: roles in mammalian development. Nat Rev Genet 2013;14(3):204–20.
[51] Bulger M, Groudine M. Functional and mechanistic diversity of distal transcription enhancers. Cell 2011;144(3):327–39.
[52] Heintzman ND, et al. Histone modifications at human enhancers reflect global cell-type-specific gene expression. Nature 2009;459(7243):108–12.
[53] Kaech SM, Ahmed R. Memory CD8+ T cell differentiation: initial antigen encounter triggers a developmental program in naive cells. Nat Immunol 2001;2(5):415–22.

[54] Gett AV, Hodgkin PD. Cell division regulates the T cell cytokine repertoire, revealing a mechanism underlying immune class regulation. Proc Natl Acad Sci U S A 1998;95(16):9488–93.
[55] Jenkins MR, et al. Cell cycle-related acquisition of cytotoxic mediators defines the progressive differentiation to effector status for virus-specific CD8+ T cells. J Immunol 2008;181(6):3818–22.
[56] Lawrence CW, Braciale TJ. Activation, differentiation, and migration of naive virus-specific CD8+ T cells during pulmonary influenza virus infection. J Immunol 2004;173(2):1209–18.
[57] Moffat JM, et al. Granzyme a expression reveals distinct cytolytic CTL subsets following influenza a virus infection. Eur J Immunol 2009;39(5):1203–10.
[58] Oehen S, Brduscha-Riem K. Differentiation of naive CTL to effector and memory CTL: correlation of effector function with phenotype and cell division. J Immunol 1998;161(10):5338–46.
[59] Denton AE, et al. Differentiation-dependent functional and epigenetic landscapes for cytokine genes in virus-specific CD8+ T cells. Proc Natl Acad Sci U S A 2011;108(37):15306–11.
[60] Wilson CB, Rowell E, Sekimata M. Epigenetic control of T-helper-cell differentiation. Nat Rev Immunol 2009;9(2):91–105.
[61] Agarwal S, Rao A. Modulation of chromatin structure regulates cytokine gene expression during T cell differentiation. Immunity 1998;9(6):765–75.
[62] Durek P, et al. Epigenomic profiling of human CD4(+) T cells supports a linear differentiation model and highlights molecular regulators of memory development. Immunity 2016;45(5):1148–61.
[63] Wei G, et al. Global mapping of H3K4me3 and H3K27me3 reveals specificity and plasticity in lineage fate determination of differentiating CD4+ T cells. Immunity 2009;30(1):155–67.
[64] Chang S, Aune TM. Dynamic changes in histone-methylation 'marks' across the locus encoding interferon-gamma during the differentiation of T helper type 2 cells. Nat Immunol 2007;8(7):723–31.
[65] Pai SY, Truitt ML, Ho IC. GATA-3 deficiency abrogates the development and maintenance of T helper type 2 cells. Proc Natl Acad Sci U S A 2004;101(7):1993–8.
[66] Zhu J, et al. Conditional deletion of Gata3 shows its essential function in T(H)1-T(H)2 responses. Nat Immunol 2004;5(11):1157–65.
[67] Schoenborn JR, et al. Comprehensive epigenetic profiling identifies multiple distal regulatory elements directing transcription of the gene encoding interferon-gamma. Nat Immunol 2007;8(7):732–42.
[68] Mullen AC, et al. Role of T-bet in commitment of TH1 cells before IL-12-dependent selection. Science 2001;292(5523):1907–10.
[69] Zhang F, Boothby M. T helper type 1-specific Brg1 recruitment and remodeling of nucleosomes positioned at the IFN-gamma promoter are Stat4 dependent. J Exp Med 2006;203(6):1493–505.
[70] Shi M, et al. Janus-kinase-3-dependent signals induce chromatin remodeling at the Ifng locus during T helper 1 cell differentiation. Immunity 2008;28(6):763–73.
[71] Hwang ES, et al. T helper cell fate specified by kinase-mediated interaction of T-bet with GATA-3. Science 2005;307(5708):430–3.
[72] Russ BE, et al. Defining the molecular blueprint that drives CD8(+) T cell differentiation in response to infection. Front Immunol 2012;3:371.
[73] Northrop JK, et al. Epigenetic remodeling of the IL-2 and IFN-gamma loci in memory CD8 T cells is influenced by CD4 T cells. J Immunol 2006;177(2):1062–9.
[74] Zediak VP, et al. Cutting edge: persistently open chromatin at effector gene loci in resting memory CD8+ T cells independent of transcriptional status. J Immunol 2011;186(5):2705–9.
[75] Araki Y, et al. Genome-wide analysis of histone methylation reveals chromatin state-based regulation of gene transcription and function of memory CD8+ T cells. Immunity 2009;30(6):912–25.
[76] Russ BE, et al. Distinct epigenetic signatures delineate transcriptional programs during virus-specific CD8(+) T cell differentiation. Immunity 2014;41(5):853–65.
[77] Russ BE, et al. Regulation of H3K4me3 at transcriptional enhancers characterizes acquisition of virus-specific CD8(+) T cell-lineage-specific function. Cell Rep 2017;21(12):3624–36.
[78] Scott-Browne JP, et al. Dynamic changes in chromatin accessibility occur in CD8+ T cells responding to viral infection. Immunity 2016;45(6):1327–40.
[79] Sen DR, et al. The epigenetic landscape of T cell exhaustion. Science 2016;354(6316):1165–9.
[80] Youngblood B, et al. Chronic virus infection enforces demethylation of the locus that encodes PD-1 in antigen-specific CD8(+) T cells. Immunity 2011;35(3):400–12.

[81] Yu B, et al. Epigenetic landscapes reveal transcription factors that regulate CD8(+) T cell differentiation. Nat Immunol 2017;18(5):573–82.
[82] Akondy RS, et al. Origin and differentiation of human memory CD8 T cells after vaccination. Nature 2017;552(7685):362–7.
[83] Bevington SL, et al. Inducible chromatin priming is associated with the establishment of immunological memory in T cells. EMBO J 2016;35(5):515–35.
[84] Lalvani A, et al. Rapid effector function in CD8+ memory T cells. J Exp Med 1997;186(6):859–65.
[85] Man K, et al. The transcription factor IRF4 is essential for TCR affinity-mediated metabolic programming and clonal expansion of T cells. Nat Immunol 2013;14(11):1155–65.
[86] Crompton JG, et al. Lineage relationship of CD8(+) T cell subsets is revealed by progressive changes in the epigenetic landscape. Cell Mol Immunol 2016;13(4):502–13.
[87] Wang D, et al. The transcription factor Runx3 establishes chromatin accessibility of cis-regulatory landscapes that drive memory cytotoxic T lymphocyte formation. Immunity 2018;48(4):659–674 e6.
[88] Dominguez CX, et al. The transcription factors ZEB2 and T-bet cooperate to program cytotoxic T cell terminal differentiation in response to LCMV viral infection. J Exp Med 2015;212(12):2041–56.
[89] Omilusik KD, et al. Transcriptional repressor ZEB2 promotes terminal differentiation of CD8+ effector and memory T cell populations during infection. J Exp Med 2015;212(12):2027–39.
[90] Kallies A, et al. Blimp-1 transcription factor is required for the differentiation of effector CD8(+) T cells and memory responses. Immunity 2009;31(2):283–95.
[91] Rutishauser RL, et al. Transcriptional repressor Blimp-1 promotes CD8(+) T cell terminal differentiation and represses the acquisition of central memory T cell properties. Immunity 2009;31(2):296–308.
[92] Pipkin ME, et al. Interleukin-2 and inflammation induce distinct transcriptional programs that promote the differentiation of effector cytolytic T cells. Immunity 2010;32(1):79–90.
[93] Nguyen ML, et al. Dynamic regulation of permissive histone modifications and GATA3 binding underpin acquisition of granzyme A expression by virus-specific CD8(+) T cells. Eur J Immunol 2016;46(2):307–18.
[94] Wang Y, et al. GATA-3 controls the maintenance and proliferation of T cells downstream of TCR and cytokine signaling. Nat Immunol 2013;14(7):714–22.
[95] Miller SA, et al. Coordinated but physically separable interaction with H3K27-demethylase and H3K4-methyltransferase activities are required for T-box protein-mediated activation of developmental gene expression. Genes Dev 2008;22(21):2980–93.
[96] Miller SA, Mohn SE, Weinmann AS. Jmjd3 and UTX play a demethylase-independent role in chromatin remodeling to regulate T-box family member-dependent gene expression. Mol Cell 2010;40(4):594–605.
[97] Sekimata M, et al. CCCTC-binding factor and the transcription factor T-bet orchestrate T helper 1 cell-specific structure and function at the interferon-gamma locus. Immunity 2009;31(4):551–64.
[98] Willinger T, et al. Human naive CD8 T cells down-regulate expression of the WNT pathway transcription factors lymphoid enhancer binding factor 1 and transcription factor 7 (T cell factor-1) following antigen encounter in vitro and in vivo. J Immunol 2006;176(3):1439–46.
[99] Kim MV, et al. The transcription factor Foxo1 controls central-memory CD8+ T cell responses to infection. Immunity 2013;39(2):286–97.
[100] Tejera MM, et al. FoxO1 controls effector-to-memory transition and maintenance of functional CD8 T cell memory. J Immunol 2013;191(1):187–99.
[101] York B, O'Malley BW. Steroid receptor coactivator (SRC) family: masters of systems biology. J Biol Chem 2010;285(50):38743–50.
[102] Shan Q, et al. The transcription factor Runx3 guards cytotoxic CD8(+) effector T cells against deviation towards follicular helper T cell lineage. Nat Immunol 2017;18(8):931–9.
[103] Arsenio J, et al. Early specification of CD8+ T lymphocyte fates during adaptive immunity revealed by single-cell gene-expression analyses. Nat Immunol 2014;15(4):365–72.
[104] Kakaradov B, et al. Early transcriptional and epigenetic regulation of CD8(+) T cell differentiation revealed by single-cell RNA sequencing. Nat Immunol 2017;18(4):422–32.
[105] Czermin B, et al. Drosophila enhancer of Zeste/ESC complexes have a histone H3 methyltransferase activity that marks chromosomal Polycomb sites. Cell 2002;111(2):185–96.

[106] Cao R, et al. Role of histone H3 lysine 27 methylation in Polycomb-group silencing. Science 2002;298(5595):1039–43.
[107] Gray SM, et al. Polycomb repressive complex 2-mediated chromatin repression guides effector CD8(+) T cell terminal differentiation and loss of multipotency. Immunity 2017;46(4):596–608.
[108] Pace L, et al. The epigenetic control of stemness in CD8(+) T cell fate commitment. Science 2018; 359(6372):177–86.
[109] Allan RS, et al. An epigenetic silencing pathway controlling T helper 2 cell lineage commitment. Nature 2012;487(7406):249–53.
[110] Agger K, et al. UTX and JMJD3 are histone H3K27 demethylases involved in HOX gene regulation and development. Nature 2007;449(7163):731–4.
[111] LaMere SA, et al. H3K27 methylation dynamics during CD4 T cell activation: regulation of JAK/ STAT and IL12RB2 expression by JMJD3. J Immunol 2017;199(9):3158–75.
[112] Youngblood B, et al. Effector CD8 T cells dedifferentiate into long-lived memory cells. Nature 2017;552(7685):404–9.
[113] Harland KL, et al. Epigenetic plasticity of Cd8a locus during CD8(+) T-cell development and effector differentiation and reprogramming. Nat Commun 2014;5:3547.
[114] Pauken KE, et al. Epigenetic stability of exhausted T cells limits durability of reinvigoration by PD-1 blockade. Science 2016;354(6316):1160–5.

This page intentionally left blank

CHAPTER 6

Microbiota in the context of epigenetics of the immune system

Katarzyna Placek
Molecular Immunology and Cell Biology Unit, Life and Medical Sciences Institute, University of Bonn, Bonn, Germany

Contents

Introduction 139
Epigenetic mechanisms 141
Gut microbiome and epigenetics of immune cells 142
  Gut microbiota and epigenetics of Treg cells 142
  Gut microbiota and epigenetics of mononuclear phagocytes 144
  Gut microbiota and epigenetics of ILCs 147
  Gut microbiota and iNKT cells 147
Skin microbiota and epigenetics of immune cells 148
Microbiome and epigenetics of nonmucosal immune cells 149
  Microbiota and nonmucosal myeloid cells 150
  The case of SCFA 150
  The case of folate 151
  Epigenetic imprint of microbes on offspring's immune cells 152
Conclusions and prospects 152
References 153

Introduction

From an integrated perspective, the role of the immune system is not restricted to fight harmful invaders but also to recognize and tolerate beneficial colonizers of the organism. Microbiota, the community of microorganisms inhabiting the body, are integrated part of the host's biology. The human body provides rich in nutrition, firm, and stable environment for microorganisms to settle, while microorganisms help to prevent from colonization by potentially dangerous microbial species and provide the host with essential nutrients for proper functioning. For example, the intestinal flora can synthetize several vitamins including folic acid (vitamin B_9), cobalamin (vitamin B_{12}), niacin (vitamin B_3), pyridoxal phosphate (the active form of vitamin B_6), pantothenic acid (vitamin B_5), biotin, tetrahydrofolate, and vitamin K. It can also affect absorption of certain minerals such as iron or can help to digest unabsorbed complex carbohydrates into short-chain fatty acids (SCFAs). This mutualistic relationship is a result of millennia of coevolution. Apart from the gastrointestinal tract, the most preferable habitats for indigenous microbiota are

Epigenetics of the Immune System
https://doi.org/10.1016/B978-0-12-817964-2.00006-X
© 2020 Elsevier Inc.
All rights reserved.

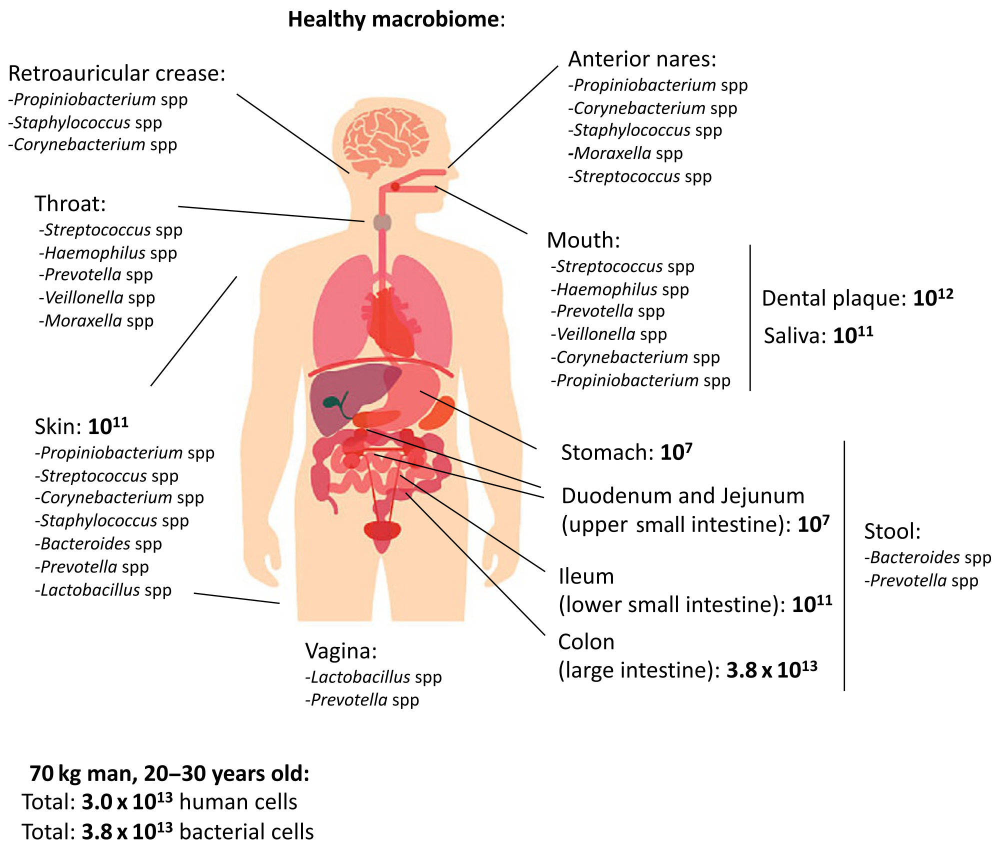

Healthy macrobiome:
Retroauricular crease:
-*Propiniobacterium* spp
-*Staphylococcus* spp
-*Corynebacterium* spp
Anterior nares:
-*Propiniobacterium* spp
-*Corynebacterium* spp
-*Staphylococcus* spp
-*Moraxella* spp
-*Streptococcus* spp
Throat:
-*Streptococcus* spp
-*Haemophilus* spp
-*Prevotella* spp
-*Veillonella* spp
-*Moraxella* spp
Mouth:
-*Streptococcus* spp
-*Haemophilus* spp
-*Prevotella* spp
-*Veillonella* spp
-*Corynebacterium* spp
-*Propiniobacterium* spp
Dental plaque: 10^{12}
Saliva: 10^{11}
Skin: 10^{11}
-*Propiniobacterium* spp
-*Streptococcus* spp
-*Corynebacterium* spp
-*Staphylococcus* spp
-*Bacteroides* spp
-*Prevotella* spp
-*Lactobacillus* spp
Stomach: 10^{7}
Duodenum and Jejunum (upper small intestine): 10^{7}
Ileum (lower small intestine): 10^{11}
Colon (large intestine): 3.8×10^{13}
Stool:
-*Bacteroides* spp
-*Prevotella* spp
Vagina:
-*Lactobacillus* spp
-*Prevotella* spp
70 kg man, 20–30 years old:
Total: 3.0×10^{13} human cells
Total: 3.8×10^{13} bacterial cells

Fig. 1 Composition and numbers of healthy human microflora at different sites of human body [1]. *(Courtesy: Katarzyna Placek.)*

oral cavity, oropharynx, vagina, and skin (Fig. 1). As a matter of fact, there are more microorganisms residing inside and on the human body than the number of cells composing the body [1]. They have an enormous impact on our health and progress of diseases and may even influence the host response to vaccines [2]. Microbiota imbalance, termed dysbiosis, has been associated with a risk of developing various disorders such as cancer, inflammatory bowel disease, celiac disease, type 1 (insulin-dependent) diabetes mellitus, obesity, chronic fatigue syndrome, bacterial vaginosis or allergy, and many more [3]. Moreover, mice reared in germ-free conditions exhibit an abnormal development of the immune system including defects in the development of gut-associated lymphoid tissues, aberrant antibody production, and reduction in size and number of Peyer's patches, and mesenteric lymph nodes [4] suggesting a direct effect of microbiota on immune cells. Indeed, postnatal colonization educates the immune system to become tolerant of a wide range of microbial immune determinants. In consequence, germ-free animals fail to develop immune tolerance and are more prone to generate allergic-like immune responses [5–10].

Raising germ-free animals and their subsequent colonization with selected microbial species or species consortium, called gnotobiotics, provide a powerful tool to look at host-microbiota interactions and the microbiota influence on the host immune system. Another powerful technology helping to address these questions is metagenomics, a study of all the genetic material recovered from an environmental sample. This method is DNA sequencing-based and culture-independent, allowing the detection of rare species and species which are not easily grown in cultures. These two approaches are nowadays the most commonly used methods in microbial research.

Epigenetic mechanisms

One of the important mechanisms by which environmental factors can influence gene expression and phenotype of an organism is epigenetic regulation. Therefore, one would expect that microbial flora can modulate the immune system development and functioning by inducing epigenetic changes in host immune cells. Epigenetic mechanisms refer to heritable changes in gene expression yet reversible and not resulting from changes in DNA sequence and include:

(1) chromatin accessibility which regulates binding of transcriptional machinery to regulatory genomic sequences [11];

(2) histone modifications which might have either silencing or activating effect on nearby genes depending on the type of modification and histone residue that they affect as well as on colocalization with other epigenetic marks; for example, histone acetylation is typically associated with transcriptional activity while histone H3 lysine K27 trimethylation (H3K27me3) is found on silenced genes [12], while H3K4 methylation is generally seen as activating modification when found together with H3K27me3 marks silenced genes but with a potential to be readily induced (called "poised" genes) [13];

(3) DNA methylation in vertebrates affects mainly CpG palindromic sequences, when found on CpG-reach promoters in general silences gene expression but its role depends on the location and a broader epigenetic context [14, 15];

(4) noncoding RNAs (ncRNAs) which contains different classes of RNAs that are not translated to proteins, could be of different sizes and execute various functions: gene silencing at posttranscriptional and posttranslational level, recruitment of DNA methylating machinery or regulation of chromatin looping (e.g., miRNA, lncRNA, etc.) [16];

(5) long-range chromatin interactions, which are responsible for bringing regulatory elements located hundreds of base pairs away from a gene to a close proximity with a promoter, therefore, enable the interaction of both and regulate gene expression. In general, more chromatin looping events happen on a locus, the more transcriptionally active is a gene within the locus [17].

All of the above play a critical role in many aspects of immune cell functioning such as immune cell development [18–21], differentiation [22, 23], activation [24–37], immune memory formation [22, 38, 39], plasticity of immune phenotypes [40–42], the maintenance of cell identity [43], and exhaustion [44–46]. Microbes can influence the host epigenome directly by producing metabolites that affect epigenetic mechanisms or indirectly by stimulating signaling pathways which results in epigenetic changes and gene expression. In this chapter, we discuss what it is known about microbiota modulation of host immune cells via their epigenomes.

Gut microbiome and epigenetics of immune cells

The most colonized human organ is the colon where 3.8×10^{13} bacteria are estimated to reside (Fig. 1) [1, 47]. Therefore, it is also the place where the most interactions between microorganisms and host cells happen. Hence intestinal flora is the most studied microbiota to date and gut microbiome, a collective genome of microorganisms inhabiting a particular environment, is the most studied microbiome nowadays. The composition of gut microflora can enforce pro- or antiinflammatory responses, among others it regulates T helper type 17 (Th17) vs regulatory (Treg) T cell balance in the lamina propria [48]. Here we will describe the effect of intestinal microbiota on epigenomes and therefore functionality of immune cell subsets.

Gut microbiota and epigenetics of Treg cells

Treg cells, which are characterized by Foxp3 gene expression, are a subset of T cells with immunosuppressive properties. Therefore, they are critical for peripheral tolerance and especially for the prevention of autoimmunity. They are also essential for immune homeostasis at barrier sites where they reassure immune tolerance to commensal microflora. Treg cells arise in the thymus and in the periphery as a consequence of exposure to microbial components such as spore-forming component, particularly clusters IV and XIVa of the bacterial class Clostridia [49], altered Schaedler flora (ASF), which is a model community of eight bacterial species of genus: *Lactobacilli*, *Bacteroides*, *Flexistipes*, and *Fusobacterium* [50], polysaccharide A of *Bacteroides fragilis* [51, 52], niacin [53], and short-chain fatty acids (SCFA) such as butyrate and propionate [53–56]. SCFAs are produced by bacteria from the phyla: Bacteroidetes and Firmicutes and are highly abundant in the large intestine [57]. Germ-free animals have significantly reduced concentrations of three main SCFAs: acetic acid, propionic acid, and butyric acid in the intestinal tract owning to a lack of enteric microbes [55, 58]. From all of the three, butyrate is the most potent immune modulator, the feature that has been linked to its histone deacetylase

(HDAC) inhibitory activity [56]. HDAC enzymes remove acetyl groups from lysine residues of histone and nonhistone proteins. The mode of action by which butyric acid blocks HDAC enzymatic activity and therefore increases acetylation levels remains unknown [59]. Histone acetylation facilitates the opening of chromatin and recruitment of transcriptional machinery as a result, it is associated with active transcription. Butyrate-treated cells have increased global levels of acetylated histones. This mechanism is believed to be responsible for the enhanced induction of Treg cells as butyrate treatment augments histone H3 acetylation in CD4$^+$ T cells at Foxp3 promoter and conserved non-coding sequences involved in Foxp3 regulation leading to a higher expression of the gene (Fig. 2) [54, 56]. Foxp3 protein itself is highly acetylated in CD4$^+$ T cells in the presence of butyrate rending the protein more resistant to proteasomal degradation [56, 60]. This observation raises the possibility that both: Foxp3 protein and histone acetylation account for enhanced Treg cell induction by HDAC inhibition and remains to be further clarified (Fig. 2). Propionate has also been shown to augment colonic Treg cell numbers and improved their IL-10 production [55]. The effect of propionate treatment on histone

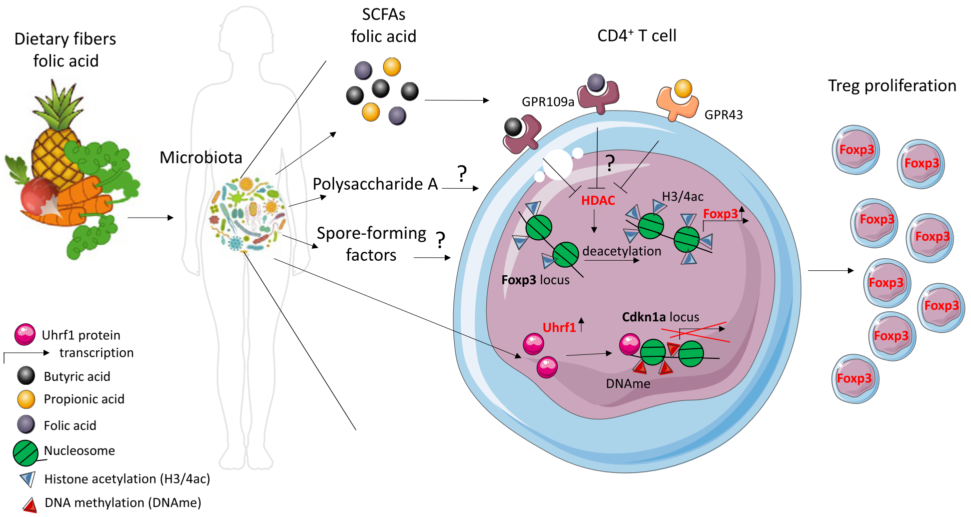

Dietary fibers
folic acid
Microbiota
SCFAs
folic acid
Polysaccharide A ?
Spore-forming factors ?
CD4+ T cell
GPR109a
GPR43
?
HDAC
H3/4ac
Foxp3↑
deacetylation
Foxp3 locus
Uhrf1↑
Cdkn1a locus
DNAme
Treg proliferation
Foxp3
Foxp3
Foxp3
Foxp3
Foxp3
Foxp3
Foxp3
Foxp3
Uhrf1 protein
transcription
Butyric acid
Propionic acid
Folic acid
Nucleosome
Histone acetylation (H3/4ac)
DNA methylation (DNAme)

Fig. 2 Epigenetic regulation of colonic Treg cell proliferation by commensal microflora. Short-chain fatty acids (SCFAs) produced by commensal microorganisms during dietary fiber fermentation are recognized by CD4$^+$ T cells via metabolite-sensing G-protein-coupled receptors (GPCRs): GPR43 and GPR109a. SCFA treatment leads to inhibition of histone deacetylases (HDACs) and accumulation of acetylated proteins, including histones at Foxp3 locus. Increased histone acetylation at regulatory elements of the gene and acetylation of Foxp3 protein leads to increased numbers of Treg cells. GPR43 can also sense niacin whose presence also induces Treg cells; however, the mechanism is still largely unknown. Commensal microflora leads to upregulation of Uhrf1 DNA methyltransferase which targets Cdkn1a gene locus. Increased DNA methylation at Cdkn1a reduces gene expression and release the blockage of cell cycle progression resulting in the proliferation of Treg cells. *(Courtesy: Katarzyna Placek.)*

acetylation in $CD4^+$ T cells seems to be less prominent than those of butyrate [54] but still significant comparing to SCFAs-untreated cells and could be a consequence of decreased expression of HDAC6 and HDAC9 in these cells [55]. SCFAs-fed mice ameliorate the development of Treg-dependent colitis indicating that Treg cells induced in the presence of SCFAs are fully functional [54, 55]. The propionate effect on chromatin modification and Treg induction is dependent on metabolite-sensing G-protein-coupled receptor (GPCR) 43 (GPR43, also called free fatty acid receptor 2 or *Ffar2*), which recognizes multiple SCFAs [61, 62]. Propionate treatment of GPR43 deficient mice does not ameliorate colitis in contrary to propionate-fed wild-type animals. A deficiency of metabolite sensor for butyrate, GPR109A, also results in reduced numbers of colonic Treg cells in lamina propria and in aggravated inflammation of the intestine [53] (Fig. 2). Defective Treg cell development in GPR109A deficient mice is partially due to a skewed ability of macrophages and dendritic cells (DCs) to promote naïve T cell differentiation toward IL-17-producing cells. This phenotype could not be restored by butyrate. Whether butyrate acts via GPR109A on chromatin has not been addressed. Interestingly, GPR109A is also a receptor for niacin, another metabolite that is synthetized by intestinal microflora. Niacin has a similar effect to butyrate on gut immune homeostasis [53]. It rescues Treg cell numbers in the intestine of antibiotic-treated mice in a GPR109A-dependent manner and suppresses colonic inflammation (Fig. 2). It would be of interest to address the niacin effect on chromatin and whether it acts via a similar signaling pathway as SCFAs.

Beside histone acetylation other epigenetic mechanism has been shown to regulate colonic Treg cell numbers and functionality by intestinal flora [63]. Uhrf1 (ubiquitin-like, with pleckstrin-homology and RING-finger domains 1) gene is highly induced in Treg cells upon colonization of sterile mice with commensal microbiota. Uhrf1 protein binds to hemi-methylated DNA and recruits DNA methyltransferase DNMT1 as well as HDAC1, therefore, forming a gene-repression complex [64–68]. Accordingly, Uhrf1 upregulation in commensal-induced Treg cells correlates with an increase in DNA methylation [63]. One of the Uhrf1 targets in Treg cells is a Cdkn1a gene (coding for cyclin-dependent kinase inhibitor p21), a cell-cycle regulator which arrests cell proliferation. Uhrf1-deficiency results in reduced DNA methylation at the distal promoter of p21 coding gene and increased gene expression. Cdkn1a silencing restores a proliferative potential of Urf1 knockout Treg cells. Thus, in commensal-induced Treg cells, Uhrf1 epigenetically silences cell-cycle inhibitor p21 allowing the robust proliferation of the cells (Fig. 2).

It is well established that mammal's colonization with commensal microbes induces Treg cells and improves the therapeutic outcome of inflammation-driven disorders. To date, epigenetic mechanisms driving this phenomenon are largely unknown.

Gut microbiota and epigenetics of mononuclear phagocytes

Mononuclear phagocytes such as macrophages and dendritic cells consist of the first line of defense against invading pathogens. They possess phagocytic and antigen-presenting properties as well as the ability to produce various cytokines. Similar to Treg cells, SCFA and especially butyrate can influence differentiation, maturation, and functionality of human monocyte-derived dendritic cells and macrophages [69–72] as well as mouse dendritic cells [56] and macrophages [73]. Dendritic cell and macrophage treatment with butyrate increases global levels of H3 acetylation and confer an antiinflammatory phenotype of these cells [56, 73]. For example, butyrate-pretreated dendritic cells enhance Treg cell differentiation from naïve CD4$^+$ T cells [56]. Moreover, stimulation of colonic lamina propria and bone marrow–derived macrophages with microbial lipopolysaccharide (LPS) in the presence of butyrate reduces IL-6, nitric oxide (NO) or nitric oxide synthase (Nos2) and IL-12p40 production but does not affect TNFα and MCP-1 chemokine levels [73]. Consistently, colonic lamina propria macrophages from antibiotic and butyrate-feed mice have reduced levels of *Il6*, *Nos2*, *Il12a*, and *Il12b* transcripts comparing to antibiotic-treated mice not exposed to butyrate. Conventional mice fed with butyrate-enriched diet show also a reduced pro-inflammatory response to *Salmonella* infection in the colon in comparison to mice fed a butyrate-free diet [72]. Interestingly, the H3K9ac levels are increased at promoters of *Nos2*, *Il6*, and *Il12β* but not *Tnfα* genes upon butyrate treatment which is not consistent with polymerase binding pattern to these regions. This is not in agreement with a general view on histone acetylation as an activating histone mark. However, it can be explained by the fact that in the absence of H3K4 methylation H3K9 acetylation facilitates the binding of the subunit of repressive Mi-2/NuRD nucleosome remodeling and deacetylase complex [74]. HDAC inhibitors enhance the DNA-binding activity of Mi-2/NuRD complex to *Il6* but not to *Tnfα* promoters in TLRs-stimulated macrophages [75] (Fig. 3). Collectively, these observations show that gene expression status is an outcome of interplay between multiple epigenetic factors. In order to fully understand molecular mechanisms controlling gene expression, one needs to look at the broad picture of epigenetic modifications as they all account for gene regulation.

The effect of commensal microflora on other epigenetic modifications is still not well understood. DCs isolated from a standard specific-pathogen-free (SPF) mice treated with DNA methylation inhibitor: 5-azacytidine produce more IFN-β, IL-6, and TNF, however, the demethylation agent did not restore cytokine production in germ-free mice [76]. This raises the question about the role of other commensal-driven epigenetic modifications in the regulation of myeloid cell activity.

Candida albicans, which is an opportunistic yeast, is found in 40%–60% of healthy humans as a component of commensal microflora of gastrointestinal tract. In immunocompromised individuals, however, it can become pathogenic and cause severe candidiasis. One of the components of the fungus recognized by immune cells is β-D-glucose polysaccharide: β-glucan. β-glucan has been shown to induce innate immune cell memory, termed "trained immunity," which can be briefly described as metabolic and epigenetic reprogramming of innate immune cells leading to an enhanced immune response toward secondary stimulus which is not related to the primary stimulus [77–79]. In in vitro and in vivo models monocyte exposure to *C. albicans* or *C. albicans*-derived

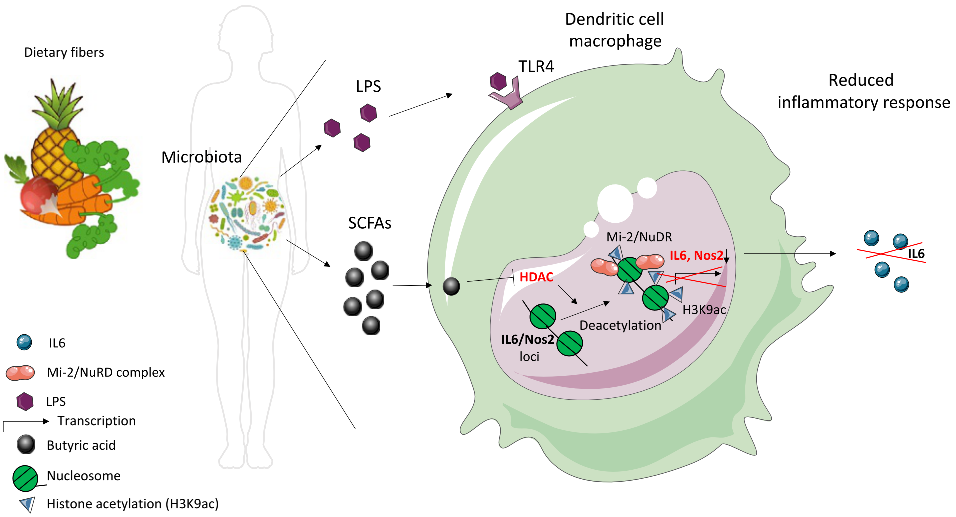

Dietary fibers
Microbiota
LPS
SCFAs
TLR4
Dendritic cell
macrophage
Reduced
inflammatory response
Mi-2/NuDR
IL6, Nos2
HDAC
H3K9ac
Deacetylation
IL6/Nos2
loci
IL6
IL6
Mi-2/NuRD complex
LPS
Transcription
Butyric acid
Nucleosome
Histone acetylation (H3K9ac)

Fig. 3 SCFAs antiinflammatory effect on dendritic cells and macrophages. SCFAs produced by microbiota inhibit HDAC activity leading to increased acetylation levels at promoter of pro-inflammatory genes such as *Il6* and *Nos2*. This leads to recruitment of Mi-2/NuDR suppressive complex to Il6 and Nos2 loci and inhibits genes expression in response to TLR4 stimulation by lipopolysaccharide (LPS). *(Courtesy: Katarzyna Placek.)*

β-glucan induces innate immune memory which is accompanied by an increase of activating histone mark H3K27 acetylation and H3K4 tri-methylation at many loci of pro-inflammatory genes including IL6, TNFα, and IL18 and other loci associated with immune response as well as multiple genes from glycolysis pathway. The latter consist of the metabolic basis of trained immunity [77, 78, 80]. Changes in H3K4me3 could be due to β-glucan-induced increase in expression of certain histone methyltransferases (e.g., SETD7). Pharmacological inhibition of histone methyltransferases using pan-methyltransferase inhibitor 5′-methylthioadenosine (MTA) compromised monocyte training indicating the histone methylation is of functional relevance in the process. Interestingly, levels of suppressive H3K27me3 do not change significantly upon exposure to β-glucan [77]. In fact, β-glucan pretreatment of monocytes leads to the appearance and disappearance of thousands of enhancers discriminated by the presence of H3K27ac, H3K4me1, and chromatin accessibility assessed by DNaseI hypersensitivity assay [78]. These changes are believed to underlay the innate immune memory phenotype and their precise role requires further functional studies. Collectively, the genome-wide assessment of chromatin structure revealed an epigenetic reprogramming of monocytes in response to β-glucan. All the above observations, however, were made for the model system without the environmental context of the intestine and the mouth where *C. albicans* is a component of healthy microflora in the majority of humans. Very likely a similar β-glucan-induced epigenetic reprogramming is happening in gastrointestinal tract but remains to be elucidated.

Gut microbiota and epigenetics of ILCs

Mucosal barriers are especially enriched in innate lymphoid cells (ILCs) which are the most recently described immune cell subset [81]. ILCs belong to lymphoid lineage but do not express antigen-specific receptors. These "helper-like" cells can be categorized into three different groups based on their transcription factors and cytokine expression profile similarity to $CD4^+$ T helper cell subsets: ILC1s which express type 1 immune program, ILC2s which express type 2 immune signature, and ILC3s which resemble Th17 cells in their transcriptional program [82]. Transcriptional signature of ILC subsets is encoded in chromatin structure. While H3K4me3 enrichment at promoters is similar between all subsets, H3K4me2 marks enhancers in ILC type-specific way supporting the hypothesis that enhancer landscape especially active enhancers (enriched in H3K27ac) is more critical in shaping cell identity than the accessibility of promoters [83]. For example, loci coding for signature transcription factors: Tbet, Gata3, and Rorc as well as signature cytokines: Ifng, Il4, and Il22 are enriched in H3K4me2 activating histone mark in ILC1, ILC2, and ILC3, respectively. Because of the high abundance of ILCs on barrier sites where most of microorganisms-host interaction happens one might expect that ILCs are highly influenced by these interactions. Interestingly mice treatment with a broad spectrum of antibiotics does not affect a general transcriptional identity of intestinal ILC1, ILC2, and ILC3 cells but affects the expression of several hundreds of genes where transcriptome of ILC1 and ILC2 are more affected than the one of ILC3 which turned out to be more stable [83]. Moreover, microbes' depletion leads to a shift of ILC1 and ILC2 transcriptional signature in a way it becomes more similar to ILC3 transcriptional program that means genes specific to ILC3 such as *Atf5*, *Gpx1* and *Cxcl9* become upregulated in ILC1 and ILC2. Similarly, the genome-wide H3K4me2 modification landscape in all ILC types are globally maintained upon antibiotic administration but a shift was observed in several thousands of H3K4me2-enriched regions especially in ILC1s. Upon treatment with antibiotics, ILC1s and ILC2s acquired H3K4me2-marked regulatory elements associated with ILC3 cell phenotype. These elements were also enriched in RORc binding motive [83]. One explanation of these findings as suggested by authors could be that the deprivation of microbial signals unleashes the default ILC3 inhibitory program. Therefore, restraining ILC3 response by microbiota might be an important mechanism in preventing ILC3-driven intestinal immune pathology [84, 85].

Gut microbiota and iNKT cells

Invariant natural killer T (iNKT) cells are a distinct population of T cells that express an invariant αβ T cell receptor (TCR) together with receptors characteristic to NK cells. Germ-free (GF) mice display increased numbers of iNKT cells in colon and lung comparing to conventionally raised specific-pathogen free (SPF) mice [86] and are more susceptible to oxazolone-induced colitis dependent on IL-13 production by

CD1c-restricted iNKT cells [87]. Similarly, sterile animals develop significantly greater allergic airway response to ovalbumin (OVA) comparing to SPF mice in an iNKT cell-dependent manner. Early colonization of mice restores normal iNKT cell numbers in lung and resistance to colitis and protects the mice in the allergic asthma model. iNKT cell accumulation in colon and lung is driven by elevated levels of *Cxcl6* gene expression in GF animals which was associated with DNA hypermethylation of CXCL6 locus. Typically DNA methylation on promoters is correlated with gene silencing. However, promoters low in CpG content can be methylated but remain active [88, 89]. One of the mechanisms behind it is the recruitment of transcriptional activator C/EBP which activates gene expression only from methylated promoters [90]. Colonization of GF neonates with a conventional microflora decreases the elevated levels of methylated DNA at CXLC6 locus. SFP neonates feeding with high doses of folic acid which increases DNA methylation levels in leukocytes [91–93] also forced DNA methylation of the *Cxcl6* gene in colon and lung and increased expression of the gene, which resulted in elevated numbers of iNKT cells. It has not been investigated which cell type in tissue samples show aberrant DNA methylation pattern at CXCL6 locus in the absence of microbiota. CXCL6 can be secreted by various cell types in response to stimuli, for example, in humans IL1-β is the predominant inducer in fibroblasts, chondrocytes, and endothelial cells [94]. Since macrophages have the capacity to produce CXCL6 especially during inflammation it is possible that interaction with commensal microbiota causes DNA methylation changes at CXCL6 locus also in these cells.

Skin microbiota and epigenetics of immune cells

Alike the gut, the skin is a habitat for complex microbial communities (Fig. 1). Commensal bacteria contribute to the skin innate immune defense by producing factors that inhibit colonization by other bacteria, termed bacteriocins, which can also interact with neutrophil extracellular traps and thus facilitate the eradication of potentially dangerous microorganisms [95]. Moreover, the role of skin flora is not restricted to provide help in fighting invaders. Recent studies have shown that epidermis-residing microbes facilitate the wound healing process by suppressing inflammation [96]. Besides, dysbiosis of the skin microflora has been linked to several skin pathologies such as acne, atopic dermatitis, psoriasis, or rosacea [95, 97]. Although little is still known about epigenetic mechanisms by which skin commensals shape local immune responses, recent reports start to unravel the role of chromatin structure in this process [98].

One of the most prevalent constituents of healthy human skin microflora is *Staphylococcus epidermidis*. The beneficial effect of *S. epidermidis* is not limited to the suppression of growth of the pathogenic *Staphylococcus aureus* as a result of the interspecies competition but also limits inflammation upon skin injury and induces CD4$^+$ T helper (Th) and CD8$^+$ T cytotoxic (Tc) cells with healing properties [98, 99]. It has been shown recently that *S. epidermidis*-specific long-lived tissue-resident memory Th17 and Tc17 cells which

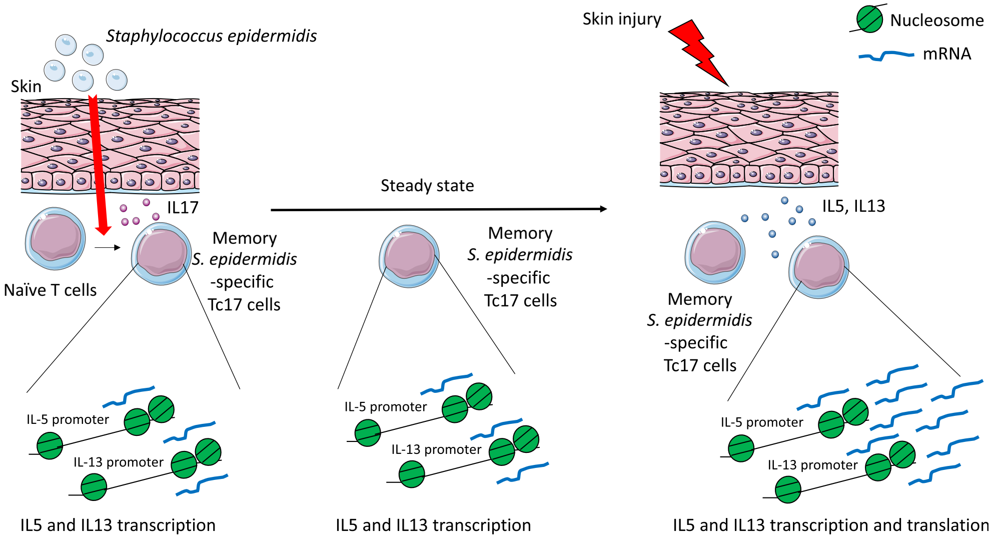

Staphylococcus epidermidis
Skin
IL17
Naïve T cells
Memory
S. epidermidis
-specific
Tc17 cells
IL-5 promoter
IL-13 promoter
IL5 and IL13 transcription
Steady state
Memory
S. epidermidis
-specific
Tc17 cells
IL-5 promoter
IL-13 promoter
IL5 and IL13 transcription
Skin injury
Nucleosome
mRNA
IL5, IL13
Memory
S. epidermidis
-specific
Tc17 cells
IL-5 promoter
IL-13 promoter
IL5 and IL13 transcription and translation

Fig. 4 Epigenetic imprint of type 2 immune response program in tissue-resident T cells by commensals. CD8$^+$ cytotoxic cell type 17 (Tc17) induced upon colonization by *S. epidermidis* possess poised type 2 immune response program which manifests by open chromatin at type 2 cytokine locus and low levels of Il5 and Il4 transcripts. Upon injury, Tc17 cells execute type 2 immune response program and contribute to the wound healing process. *(Courtesy: Katarzyna Placek.)*

normally express RORgt transcription factor and IL17a cytokine are able to produce high amounts of IL-5 and IL-13 type 2 signature cytokines upon injury. The IL-13 production by the commensal-specific Tc17 promotes wound repair. The ability to produce type 2 cytokines by Tc17 is enabled by the presence of low levels of Il13 and Il5 transcripts in these cells as well as a Tc17-specific chromatin structure induced during *S. epidermidis* colonization which is characterized by the accessibility of Il5 and Il13 promoters [98] (Fig. 4). This is the first report showing that commensal microorganisms can induce type 17 T cells with a potential of type 2 response and this transcriptional plasticity is encoded in chromatin structure.

Similar to the gut microbiota, a pathogenic *Propionibacterium acnes* produce SCFAs, which inhibit the activity of histone deacetylases in neighbor keratinocytes leading to gene activation and release of pro-inflammatory cytokines [100]. Most likely, although not described so far, they may have a similar effect on epigenetic regulation of skin-resident immune cells.

Microbiome and epigenetics of nonmucosal immune cells

Microbial colonization has its imprint on the epigenome of multiple host tissues [101]. Epigenetic regulation of intestinal epithelial cells by aberrant gut microbiota has been well appreciated as a mechanism leading to, for example, colon cancer development

[102]. Direct contact of microorganism with the host's cell is not always necessary for having an effect on the epigenome. Immune cells that are not residing at barrier sites also succumb to the influence of commensal colonization.

Microbiota and nonmucosal myeloid cells

Sterile or antibiotic-treated mice display normal numbers of splenic mononuclear phagocytes. However, nonmucosal dendritic cells and macrophages from germ-free animals fail to upregulate various inflammatory response genes such as interferons type I (IFN-I), Il15, IL-15Ra, Il6, Tnf, Il12, and Il18 upon exposure to microbial ligands such as poly(I:C) and LPS or mouse cytomegalovirus (MCMV) infection [76]. In consequence, the IFN-I response genes such as *Ifr7* and *Cxcl10*, which consist a part of the transcriptional program required for viral defense, were not upregulated upon microbial stimulation. The defect in IFN-I expression in activated DCs is cell-intrinsic, however, it does not result from abrogated phosphorylation of IRF3 and IkBa which mediates pattern recognition receptor signaling but the ability of these factors to bind to target DNA. This points to the chromatin structure at target genes being in a permissive conformation in germ-free mice preventing the binding of transcriptional machinery. Indeed, deprivation of commensal microflora resulted in lower levels of activating histone mark H3K4me3 at promoters of microbiota-dependent inflammatory genes (e.g., *Tnfα* and *Il6*) in DCs. The data indicate that signals from the commensal microbiota may induce deposition of activating histone marks poising the pro-inflammatory cytokine genes for expression in myeloid cells. However, HDAC inhibitors (sodium butyrate and suberoylanilide hydroxamic acid: SAHA) treatment of DCs isolated from germ-free mice did not restore cytokine production upon microbial challenge [76].

The case of SCFA

SCFA production by microbiota is the most prominent example of microbial flora direct effect on immune cell epigenome. In addition to in vivo studies which revealed the epigenome modulating properties on colonic immune cells, the in vitro studies on blood-derived immune cells also show the effect of SCFA on peripheral immune cells especially macrophages, dendritic cells, and neutrophils where SCFAs have been shown to inhibit the production of inflammatory cytokines by these cells or improve their bactericidal function [72, 73, 103, 104]. Macrophages differentiated in the presence of butyrate and propionate but not acetate display enhanced antimicrobial activity [72]. This effect could be mimic by the use of other pan-HDAC inhibitors such as phenyl-butyrate and veronistat (SAHA) and specific HDAC class I inhibitors: valproate (targets HDAC1-3), SBHA (targets HDAC1 and 3), and RGFP966 (targets HDAC3) but not by specific HDAC class II inhibitors: TMP195 (HDAC class IIa) and tubacin (HDAC class IIb) nor HDAC class I inhibitor 1-naphthohydroxamic acid (NA, targets HDAC8, 1, and 6).

RNA interference experiments confirmed that HDAC3 silencing is sufficient to increase the antimicrobial activity of macrophages and butyrate had no additional effect in the presence of the siRNA targeting HDAC3. This suggests that butyrate act via HDAC3 inhibition in blood-derived macrophages. As expected butyrate-treatment leads to increased global levels of H3 acetylation, interestingly it also causes a reduction in suppressive histone mark: H3K27me3. However, the HDAC3 targets in macrophage remain undescribed. Butyrate-induced improvement of macrophage functionality does not require GPRC receptor as the addition of GPCR inhibitor: pertussis toxin (PT) to the culture does not reverse the effect of butyrate.

Even though plasma and serum levels of SCFAs in human blood are typically very low [105] the SCFA effect on peripheral blood mononuclear immune cells observed in vitro might be still a systemic effect and of relevance in vivo.

The case of folate

Folic acid, known as vitamin B_9, is a critical micronutrient which can be absorbed from specific foods and which is synthetized by commensal bacteria *Bifidobacterium* and *Lactobacillus* [106, 107]. As a coenzyme of one-carbon metabolism, it reinforces various physiological processes including methyl group transfer for DNA methylation. Moderate folate depletion causes a decrease in DNA methylation levels in leukocytes [91] while folate supplementation in the diet increases global DNA methylation levels in leukocytes in peripheral blood [91, 93] with one region of 932 bp 3 kb upstream of the *ZFP57* gene being particularly affected [92]. Moreover, maternal folate concentrations during pregnancy have a systemic effect on DNA methylation status on cord blood [108] and in the offspring in a tissue-dependent way [109, 110] and have been associated with many numerous defects including neural tube defects [109], childhood wheeze [111], atopic dermatitis [33, 112], or asthma [113] and others. Genome-wide DNA methylation assessment of adaptive ($CD4^+$ T cells) and innate ($CD14^+$ antigen-presenting cells, APCs) immune cells isolated from neonatal cord blood of mothers with high and low levels of folic acid revealed several differentially methylated regions between two groups from which a 932 bp region 3 kb upstream of the *ZFP57* gene was hypomethylated in both cell types in a group of mothers with high folate concentrations [114]. Decrease of DNA methylation of the region in high folate condition was accompanied by increased H3 and H4 acetylation levels and in consequence gene expression suggesting a functional role of these epigenetic differences on gene expression. These data were not confirmed, however, in neonatal cord blood samples of mothers who underwent folate supplementation during pregnancy where the opposite pattern of DNA methylation of ZFP57 was found [92]. Even though contradictory the studies point to the ZFP57 gene being folate-sensitive. ZFP57 is thought to maintain the methylation status of a cell. What is a consequence of the differential ZFP57 transcript levels on immune cells remains unanswered.

These observations suggest that folate production by gut microflora could also modulate the epigenome of immune cells. In fact, the administration of probiotic *Bifidobacterium* strains increases folate concentrations in human feces [107] and animal tissues [115]. Very likely folate-producing microbes modulate DNA methylation landscape of immune cells. Yet, this effect needs to be assessed especially the functional relevance of folate-induced DNA methylation increase on immune cell biology.

Epigenetic imprint of microbes on offspring's immune cells

Increasing prevalence of allergic responses and asthma in children in developed countries can be explained by "hygiene hypothesis" which states that lack of exposure to pathogenic and symbiotic microorganisms in early childhood increases susceptibility to allergic diseases and immune intolerance by abrogating a natural development of the immune system. In accordance with this hypothesis, multiple reports have linked gut dysbiosis with the development of allergic type immune responses [116]. Moreover, maternal exposure to microorganisms during gestation has its effect on offspring immune system. For example, pups born from mothers prenatally supplemented with *Lactobacillus rhamnosus GG* display reduced cytokine production such as TNFα, IFNγ, IL10, and IL5 but not IL4 nor IL13 [117]. Living on a farm especially in early childhood and in prenatal period has been also linked to protection against hay fever, asthma, and allergic sensitization [118–122]. The main constituent of the farm environment that is thought to drive the protection by modulating immune system are microbial components especially endotoxins and the Gram-negative, nonpathogenic bacterium *Acinetobacter lwoffii* F78 and *Lactoccocus lactis* G121 [123, 124]. Maternal exposure to *A. lwoffii* F78 during pregnancy confers protection from the development of asthmatic phenotype in offspring which was manifested by reduced production of type 2 cytokines but not IFNγ [125]. Cytokine expression pattern in CD4$^+$ T cells of progeny corresponds to H4 acetylation levels at cytokine gene promoters while suppressive histone mark H3K27me3 levels were not affected by *A. lwoffii* F78 pretreatment and they were barely detectable at cytokine gene promoters. Consistently pharmacological inhibition of histone de novo acetylation process abrogated a protective effect of *A. lwoffii* F78. Thus, transmaternal protection against asthma is mediated via epigenetic changes that affect cytokine loci in CD4$^+$ T cells in progeny.

Conclusions and prospects

The appreciation of the effect of the microbiome on human health and development is growing exponentially. Probiotics are one of the most commonly consumed food supplements and the probiotic industry is constantly expanding. There is a multitude of suggested prophylactic and therapeutic indications such as amelioration of inflammatory bowel disease and irritable bowel syndrome, treatment of acute diarrheas or reduction of risk for neonatal late-onset sepsis and necrotizing enterocolitis, eradication of *Helicobacter pylori,* prevention and treatment of atopic dermatitis and many others for probiotic use. Yet, still controversial in their efficacy microbiota-based therapies are a promising tool in fighting many types of pathologies [126]. As we have learned a lot about the modulation of the human immune system by microbiota only scare of information we know about the epigenetic imprint of different microorganisms on immune cells. While the noncoding RNA expression profile of gut reflects the composition of intestinal flora [127] it would be interesting to know whether tissue-resident immune cells possess a similar commensal-specific ncRNA transcriptome signature and how does it affect immune function. Likewise, nothing is known about chromatin interactions in immune cells upon microbial exposure and whether DNA looping has also its role in transmitting the microflora effect on immune cell phenotype. Even though the microbiome-driven immunomodulation is a long-lasting concept we are only starting to unravel epigenetic mechanisms behind it. And as epigenetic modulators are becoming more and more in use for the treatment of various pathologies including disorders of the immune system they might be also a promising cure for disorders resulting from dysbiosis. In order to elucidate their therapeutic potential deeper understanding of the effect of microbiome on the epigenetics of the host immune system is required.

References

[1] Sender R, Fuchs S, Milo R. Revised estimates for the number of human and bacteria cells in the body. PLoS Biol 2016;14(8). e1002533.

[2] Macpherson AJ. Do the microbiota influence vaccines and protective immunity to pathogens? Cold Spring Harb Perspect Biol 2018;10(2):a029363.

[3] Tlaskalova-Hogenova H, Stepankova R, Kozakova H, Hudcovic T, Vannucci L, Tuckova L, et al. The role of gut microbiota (commensal bacteria) and the mucosal barrier in the pathogenesis of inflammatory and autoimmune diseases and cancer: contribution of germ-free and gnotobiotic animal models of human diseases. Cell Mol Immunol 2011;(December 2010):110–20.

[4] Shanahan F. The host ± microbe interface within the gut. Best Pract Res Clin Gastroenterol 2002; 16(6):915–31.

[5] Sudo N, Sawamura S, Tanaka K, Aiba Y, Kubo C, Koga Y. The requirement of intestinal bacterial flora for the development of an IgE production system fully susceptible to oral tolerance induction. J Immunol 1997;159(4):1739–45.

[6] Braun-Fahrlander C, Riedler J, Herz U, Eder W, Waser M, Grize L, et al. Numb Er 12 environmental exposure to endotoxin and its relation. N Engl J Med 2002;347(12):869–77.

[7] Herbst T, Sichelstiel A, Scha C, Yadava K, Burki K, Cahenzli J, et al. Dysregulation of allergic airway inflammation in the absence of microbial colonization. Am J Respir Crit Care Med 2011;184(2): 198–205.

[8] Rodriguez B, Prioult G, Bibiloni R, Nicolis I, Mercenier A, Butel M-J, et al. Germ-free status and altered caecal subdominant microbiota are associated with a high susceptibility to cow's milk allergy in mice. FEMS Microbiol Ecol 2011;76:133–44.

[9] Stefka AT, Feehley T, Tripathi P, Qiu J, Mccoy K, Mazmanian SK. Commensal bacteria protect against food allergen sensitization. Proc Natl Acad Sci USA 2014;111(36):2–7.

[10] Schwarzer M, Srutkova D, Hermanova P, Leulier F, Kozakova H. Schabussova I. Diet matters: endotoxin in the diet impacts the level of allergic sensitization in germ-free mice 2017;1–15.

[11] Klemm SL, Shipony Z, Greenleaf WJ. Chromatin accessibility and the regulatory epigenome. Nat Rev Genet 2019;20(April):29–35.

[12] Bannister AJ, Kouzarides T. Regulation of chromatin by histone modifications. Nat Publ Gr 2011;21 (3):381–95.

[13] Bernstein BE, Mikkelsen TS, Xie X, Kamal M, Huebert DJ, Cuff J, et al. A bivalent chromatin structure marks key developmental genes in embryonic stem cells. Cell 2006;315–26.

[14] Schübeler D. Function and information content of DNA methylation. Nature 2015;517:321–6.

[15] Luo C, Hajkova P, Ecker JR. Dynamic DNA methylation: in the right place at the right time. Science 2018;1340(September):1336–40.

[16] Holoch D, Moazed D. RNA-mediated epigenetic regulation of gene expression. Nat Publ Gr 2015;16(2):71–84.

[17] Dekker J, Misteli T. Long-range chromatin interactions. Cold Spring Harb Perspect Biol 2015;7(10), a019356.

[18] Hu G, Cui K, Fang D, Hirose S, Wang X, Wangsa D, et al. Transformation of accessible chromatin and 3D nucleome underlies lineage commitment of early T cells. Immunity 2018;48(2):227–42.

[19] Johnson JL, Georgakilas G, Petrovic J, Kurachi M, Cai S, Harly C, et al. Lineage-determining transcription factor TCF-1 initiates the epigenetic identity of T cells. Immunity 2018;48(2):243–57.

[20] Mandal M, Maienschein-Cline M, Maffucci P, Veselits M, Kennedy DE, McLean KC, et al. BRWD1 orchestrates epigenetic landscape of late B lymphopoiesis. Nat Commun 2018;9(1).

[21] Hewitt SL, Chaumeil J, Skok JA. Chromosome dynamics and the regulation of V(D)J recombination. Immunol Rev 2010;237(1):43–54.

[22] Álvarez-Errico D, Vento-Tormo R, Sieweke M, Ballestar E. Epigenetic control of myeloid cell differentiation, identity and function. Nat Rev Immunol 2015;15(1):7–17.

[23] Schmidl C, Delacher M, Huehn J, Feuerer M. Epigenetic mechanisms regulating T-cell responses. J Allergy Clin Immunol 2018;142(3):728–43.

[24] Agarwal S, Rao A. Modulation of chromatin structure regulates cytokine gene expression during T cell differentiation. Immunity 1998;9(6):765–75.

[25] Eivazova ER, Aune TM. Dynamic alterations in the conformation of the Ifng gene region during T helper cell differentiation. Proc Natl Acad Sci 2004;101(1):251–6.

[26] Lee GR, Spilianakis CG, Flavell RA. Hypersensitive site 7 of the TH2 locus control region is essential for expressing TH2 cytokine genes and for long-range intrachromosomal interactions. Nat Immunol 2005;6(1):42–8.

[27] Tsytsykova AV, Rajsbaum R, Falvo JV, Ligeiro F, Neely SR, Goldfeld AE. Activation-dependent intrachromosomal interactions formed by the TNF gene promoter and two distal enhancers. Proc Natl Acad Sci 2007;104(43):16850–5.

[28] Park J-H, Choi Y, Song M-J, Park K, Lee J-J, Kim H-P. Dynamic long-range chromatin interaction controls expression of IL-21 in CD4+ T Cells. J Immunol 2016;196(10):4378–89.

[29] Sharaf N, Nicklin MJ, Di Giovine FS. Long-range DNA interactions at the IL-1/IL-36/IL-37 gene cluster (2q13) are induced by activation of monocytes. Cytokine 2014;68(1):16–22.

[30] Placek K, Gasparian S, Coffre M, Maiella S, Sechet E, Bianchi E, et al. Integration of distinct intracellular signaling pathways at distal regulatory elements directs T-bet expression in human CD4+ T cells. J Immunol 2009;183(12):7743–51.

[31] Bevington SL, Cauchy P, Piper J, Bertrand E, Lalli N, Jarvis RC, et al. Inducible chromatin priming is associated with the establishment of immunological memory in T cells. EMBO J 2016;35(5):515–35.

[32] Scott-Browne JP, López-Moyado IF, Trifari S, Wong V, Chavez L, Rao A, et al. Dynamic changes in chromatin accessibility occur in CD8 + T cells responding to viral infection. Immunity 2016;45(6): 1327–40.

[33] Kieffer-Kwon KR, Nimura K, Rao SSP, Xu J, Jung S, Pekowska A, et al. Myc regulates chromatin decompaction and nuclear architecture during B cell activation. Mol Cell 2017;67(4). 566-578.e10.

[34] Phan AT, Goldrath AW, Glass CK. Metabolic and epigenetic coordination of T cell and macrophage immunity. Immunity 2017;46(5):714–29.

[35] Iiott NE, Heward JA, Roux B, Tsitsiou E, Fenwick PS, Lenzi L, et al. Long non-coding RNAs and enhancer RNAs regulate the lipopolysaccharide-induced inflammatory response in human monocytes. Nat Commun 2014;5.

[36] Carpenter S, Aiello D, Atianand MK, Ricci EP, Gandhi P, Hall LL, et al. A long noncoding RNA mediates both activation and repression of immune response genes. Science 2014;341(August 2013):789–92.

[37] Cui H, Xie N, Tan Z, Banerjee S, Thannickal VJ, Abraham E, et al. The human long noncoding RNA lnc-IL7R regulates the inflammatory response. Eur J Immunol 2014;44(7):2085–95.

[38] Netea MG, Schlitzer A, Placek K, Joosten LAB, Schultze JL. Innate and adaptive immune memory: an evolutionary continuum in the Host's response to pathogens. Cell Host Microbe 2019;25(1): 13–26.

[39] Barski A, Cuddapah S, Kartashov AV, Liu C, Imamichi H, Yang W, et al. Rapid recall ability of memory T cells is encoded in their epigenome. Sci Rep 2017;7(November 2016):1–10.

[40] He H, Ni B, Tian Y, Tian Z, Chen Y, Liu Z, et al. Histone methylation mediates plasticity of human FOXP3 + regulatory T cells by modulating signature gene expressions. Immunology 2014;141(3): 362–76.

[41] Cheray M, Joseph B. Epigenetics control microglia plasticity. Front Cell Neurosci 2018;12 (August):1–13.

[42] Satoh T, Tajima M, Wakita D, Kitamura H, Nishimura T. The development of IL-17/IFN-γ-double producing CTLs from Tc17 cells is driven by epigenetic suppression of Socs3 gene promoter. Eur J Immunol 2012;42(9):2329–42.

[43] Polansky JK, Kretschmer K, Freyer J, Floess S, Garbe A, Baron U, et al. DNA methylation controls Foxp3 gene expression. Eur J Immunol 2008;38(6):1654–63.

[44] McLane LM, Abdel-Hakeem MS, Wherry EJ. CD8 T cell exhaustion during chronic viral infection and cancer. Annu Rev Immunol 2019;37(1):457–95.

[45] Wherry EJ, Godec J, Haining WN, Berger SL, Bartman C, Sen DR, et al. Epigenetic stability of exhausted T cells limits durability of reinvigoration by PD-1 blockade. Science 2016;354 (6316):1160–5.

[46] Sen DR, Kaminski J, Barnitz RA, Kurachi M, Gerdemann U, Yate KB, et al. The epigenetic landscape of T cell exhaustion. Science 2016;0491(October):1–6.

[47] Sender R, Fuchs S, Milo R. Are we really vastly outnumbered? Revisiting the ratio of bacterial to host cells in humans. Cell 2016;164(3):337–40.

[48] Ivanov II, Frutos Rde L, Manel N, Yoshinaga K, Rifkin DB, Sartor RB, et al. Specific microbiota direct the differentiation of IL-17-producing T-helper cells in the mucosa of the small intestine. Cell Host Microbe 2008;4(4):337–49.

[49] Atarashi K, Tanoue T, Shima T, Imaoka A, Kuwahara T, Momose Y, et al. Induction of colonic regulatory T cells by indigenous Clostridium species. Science 2011;331(January):337–41.

[50] Geuking MB, Cahenzli J, Lawson MAE, Ng DCK, Slack E, Hapfelmeier S, et al. Intestinal bacterial colonization induces mutualistic regulatory T cell responses. Immunity 2011;34(5):794–806.

[51] Round JL, Mazmanian SK. Inducible Foxp3 + regulatory T-cell development by a commensal bacterium of the intestinal microbiota. PNAS 2010;107(27):12204–9.

[52] Round JL, Lee SM, Li J, Tran G, Jabri B, Chatila TA, et al. The toll-like receptor 2 pathway establishes colonization by a commensal of the human microbiota. Science 2011;332(6032):974–7.

[53] Singh N, Gurav A, Sivaprakasam S, Brady E, Padia R, Shi H, et al. Activation of Gpr109a, receptor for niacin and the commensal metabolite butyrate, suppresses colonic inflammation and carcinogenesis. Immunity 2014;40(1):128–39.

[54] Furusawa Y, Obata Y, Fukuda S, Endo TA, Nakato G, Takahashi D, et al. Commensal microbe-derived butyrate induces the differentiation of colonic regulatory T cells. Nature 2013;504 (7480):446–50.

[55] Smith PM, Howitt MR, Panikov N, Michaud M, Gallini CA, Bohlooly-y M, et al. The microbial metabolites, short-chain fatty acids, regulate colonic treg cell homeostasis. Science 2013;341 (August):569–74.

[56] Arpaia N, Campbell C, Fan X, Dikiy S, Van Der Veeken J, Deroos P, et al. Metabolites produced by commensal bacteria promote peripheral regulatory T-cell generation. Nature 2013;504(7480):451–5.

[57] Louis P, Flint HJ. Diversity, metabolism and microbial ecology of butyrate-producing bacteria from the human large intestine. FEMS Microbiol Lett 2009;294(1):1–8.

[58] Hoverstad T, Midtvedt T. Short-chain fatty acids in germfree mice and rats. J Nutr 1986;116(9):1772–6.
[59] Petricoin EF, Liotta LA. Nutritional proteomics in cancer prevention clinical applications of proteomics 1. J Nutr 2003;133:2476–84.
[60] Van Loosdregt J, Vercoulen Y, Guichelaar T, Gent YYJ, Beekman JM, Van Beekum O, et al. Regulation of Treg functionality by acetylation-mediated Foxp3 protein stabilization. Blood 2010;115(5):965–74.
[61] Brown AJ, Goldsworthy SM, Barnes AA, Eilert MM, Tcheang L, Daniels D, et al. The orphan G protein-coupled receptors GPR41 and GPR43 are activated by propionate and other short chain carboxylic acids. J Biol Chem 2003;278(13):11312–9.
[62] Le Poul E, Loison C, Struyf S, Springael JY, Lannoy V, Decobecq ME, et al. Functional characterization of human receptors for short chain fatty acids and their role in polymorphonuclear cell activation. J Biol Chem 2003;278(28):25481–9.
[63] Obata Y, Furusawa Y, Endo TA, Sharif J, Takahashi D, Atarashi K, et al. The epigenetic regulator Uhrf1 facilitates the proliferation and maturation of colonic regulatory T cells. Nat Immunol 2014;15(6):571–9.
[64] Bostick M, Kim JK, Esteve P-O, Clark A, Pradhan S, Jackobsen SE. UHRF1 plays a role in maintaining DNA methylation in mammalian cells. Science (80-) 2007;317(September):1760–5.
[65] Sharif J, Muto M, Takebayashi SI, Suetake I, Iwamatsu A, Endo TA, et al. The SRA protein Np95 mediates epigenetic inheritance by recruiting Dnmt1 to methylated DNA. Nature 2007;450(7171):908–12.
[66] Unoki M, Nishidate T, Nakamura Y. ICBP90, an E2F-1 target, recruits HDAC1 and binds to methyl-CpG through its SRA domain. Oncogene 2004;23(46):7601–10.
[67] Rezasoltani S, Asadzadeh-Aghdaei H, Nazemalhosseini-Mojarad E, Dabiri H, Ghanbari R, Zali MR. Gut microbiota, epigenetic modification and colorectal cancer. Iran J Microbiol 2017;9(2):55–63.
[68] Nishiyama A, Yamaguchi L, Sharif J, Johmura Y, Kawamura T, Nakanishi K, et al. Uhrf1-dependent H3K23 ubiquitylation couples maintenance DNA methylation and replication. Nature 2013;502(7470):249–53.
[69] Millard AL, Mertes PM, Ittelet D, Villard F, Jeannesson P, Bernard J. Butyrate affects differentiation, maturation and function of human monocyte-derived dendritic cells and macrophages. Clin Exp Immunol 2002;130(2):245–55.
[70] Wang B, Morinobu A, Horiuchi M, Liu J, Kumagai S. Butyrate inhibits functional differentiation of human monocyte-derived dendritic cells. Cell Immunol 2008;253(1–2):54–8.
[71] Nastasi C, Candela M, Bonefeld CM, Geisler C, Hansen M, Krejsgaard T, et al. The effect of short-chain fatty acids on human monocyte-derived dendritic cells. Sci Rep 2015;5:1–10.
[72] Schulthess J, Pandey S, Capitani M, Rue-Albrecht KC, Arnold I, Franchini F, et al. The short chain fatty acid butyrate imprints an antimicrobial program in macrophages. Immunity 2019;50(2). 432-445.e7.
[73] Chang PV, Hao L, Offermanns S, Medzhitov R. The microbial metabolite butyrate regulates intestinal macrophage function via histone deacetylase inhibition. Proc Natl Acad Sci 2014;111(6):2247–52.
[74] Musselman C, Mansfield R, Garske AL, Davrazou F, Kwan AH, Oliver SS, et al. Binding of the CHD4 PHD2 finger to histone H3 is modulated by covalent modifications. Biochem J 2010;423(2):179–87.
[75] Roger T, Le Roy D, Mombelli M, Koessler T, Ding XC, Chanson A, et al. Histone deacetylase inhibitors impair innate immune responses to Toll-like receptor agonists and to infection. Blood 2011;117(4):1205–17.
[76] Ganal SC, Sanos SL, Kallfass C, Oberle K, Johner C, Kirschning C, et al. Priming of natural killer cells by nonmucosal mononuclear phagocytes requires instructive signals from commensal microbiota. Immunity 2012;37(1):171–86.
[77] Quintin J, Saeed S, Martens JHA, Giamarellos-Bourboulis EJ, Ifrim DC, Logie C, et al. Candida albicans infection affords protection against reinfection via functional reprogramming of monocytes. Cell Host Microbe 2012;12(2):223–32.

[78] Saeed S, Quintin J, Kerstens HHD, Rao NA, Aghajanirefah A, Matarese F, et al. Epigenetic programming of monocyte-to-macrophage differentiation and trained innate immunity. Science (80-) 2014;345(6204).

[79] Netea MG, Joosten LAB, Latz E, Mills KHG, Natoli G, Stunnenberg HG, et al. Trained immunity: a program of innate immune memory in health and disease. Science 2016;352(6284).

[80] Cheng SC, Quintin J, Cramer RA, Shepardson KM, Saeed S, Kumar V, et al. MTOR- and HIF-1α-mediated aerobic glycolysis as metabolic basis for trained immunity. Science 2014;345(6204).

[81] Eberl G, Colonna M, Santo JPD, ANJ MK. Innate lymphoid cells: a new paradigm in immunology. Science 2015;348(6237).

[82] Serafini N, Vosshenrich CAJ, Di Santo JP, Serafini N, Vosshenrich CAJ, Di JP, et al. Transcriptional regulation of innate lymphoid cell fate. Nat Rev Immunol 2015;15(7):415–28.

[83] Gury-BenAri M, Thaiss CA, Serafini N, Winter DR, Giladi A, Lara-Astiaso D, et al. The spectrum and regulatory landscape of intestinal innate lymphoid cells are shaped by the microbiome. Cell 2016;166(5). 1231-1246.e13.

[84] Buonocore S, Ahern PP, Uhlig HH, Ivanov II, Littman DR, Maloy KJ, et al. Innate lymphoid cells drive interleukin-23-dependent innate intestinal pathology. Nature 2010;464(7293):1371–5.

[85] Balzola F, Bernstein C, Ho GT, Russell RK. IL-23-responsive innate lymphoid cells are increased in inflammatory bowel disease: commentary. Inflamm Bowel Dis Monit 2011;12(2):74.

[86] Olszak T, An D, Zeissig S, Vera MP, Richter J, Andre Franke JNG, et al. Microbial exposure during early life has persistent effects on natural killer T cell function. Science 2012;336(6080):489–93.

[87] Heller F, Fuss IJ, Nieuwenhuis EE, Blumberg RS, Strober W. Oxazolone colitis, a Th2 colitis model resembling ulcerative colitis, is mediated by IL-13-producing NK-T cells. Immunity 2002;17(5): 629–38.

[88] Eckhardt F, Lewin J, Cortese R, Rakyan VK, Attwood J, Burger M, et al. DNA methylation profiling of human chromosomes 6, 20 and 22. Nat Genet 2006;38(12):1378–85.

[89] Weber M, Hellmann I, Stadler MB, Ramos L, Pääbo S, Rebhan M, et al. Distribution, silencing potential and evolutionary impact of promoter DNA methylation in the human genome. Nat Genet 2007;39(4):457–66.

[90] Rishi V, Bhattacharya P, Chatterjee R, Rozenberg J, Zhao J, Glass K, et al. CpG methylation of half-CRE sequences creates C/EBP binding sites that activate some tissue-specific genes. Proc Natl Acad Sci 2010;107(47):20311–6.

[91] Jacob RA, Gretz DM, Taylor PC, James SJ, Pogribny IP, Miller BJ, et al. Moderate folate depletion increases plasma homocysteine and decreases lymphocyte DNA methylation in postmenopausal women 1–4. J Nutr 1998;128(January):1204–12.

[92] Irwin RE, Thursby SJ, Ondičová M, Pentieva K, McNulty H, Richmond RC, et al. A randomized controlled trial of folic acid intervention in pregnancy highlights a putative methylation-regulated control element at ZFP57. Clin Epigenetics 2019;11(1):1–16.

[93] Pufulete M, Al-Ghnaniem R, Khushal A, Appleby P, Harris N, Gout S, et al. Effect of folic acid supplementation on genomic DNA methylation in patients with colorectal adenoma. Gut 2005;54(5): 648–53.

[94] Wuyts A, Struyf S, Gijsbers K, Schutyser E, Put W, Conings R, et al. The CXC chemokine GCP-2/CXCL6 is predominantly induced in mesenchymal cells by interleukin-1β and is down-regulated by interferon-γ: comparison with interleukin-8/CXCL8. Lab Investig 2003;83(1):23–34.

[95] Gallo RL, Nakatsuji T. Microbial symbiosis with the innate immune defense system of the skin. J Invest Dermatol 2011;131(10):1974–80.

[96] Lai Y, Di Nardo A, Nakatsuji T, Leichtle A, Yang Y, Cogen AL, et al. Commensal bacteria regulate toll-like receptor 3-dependent inflammation after skin injury. Nat Med 2009;15(12):1377–82.

[97] Prescott SL, Larcombe DL, Logan AC, West C, Burks W, Caraballo L, et al. The skin microbiome: Impact of modern environments on skin ecology, barrier integrity, and systemic immune programming. World Allergy Organ J 2017;10(1):1–16.

[98] Harrison OJ, Linehan JL, Shih H-Y, Bouladoux N, Han S-J, Smelkinson M, et al. Commensal-specific T cell plasticity promotes rapid tissue adaptation to injury. Science 2019;363(6422):eaat6280.

[99] Stacy A, Belkaid Y. Microbial guardians of skin health. Science 2019;363(6424):227–8.

[100] Sanford JA, Zhang LJ, Williams MR, Gangoiti JA, Huang CM, Gallo RL. Inhibition of HDAC8 and HDAC9 by microbial short-chain fatty acids breaks immune tolerance of the epidermis to TLR ligands. Sci Immunol 2016;1(4).

[101] Krautkramer KA, Kreznar JH, Romano KA, Attie AD, Rey FE, Denu JM, et al. Short article: diet-microbiota interactions mediate global epigenetic programming in multiple host tissues. Mol Cell 2016;64(5):982–92.

[102] Yang T, Owen JL, Lightfoot YL, Kladde MP, Mohamadzadeh M. Microbiota impact on the epigenetic regulation of colorectal cancer. Trends Mol Med 2013;19(12):714–25.

[103] Usami M, Kishimoto K, Ohata A, Miyoshi M, Aoyama M, Fueda Y, et al. Butyrate and trichostatin A attenuate nuclear factor κB activation and tumor necrosis factor α secretion and increase prostaglandin E2 secretion in human peripheral blood mononuclear cells. Nutr Res 2008;28(5):321–8.

[104] Vinolo MAR, Rodrigues HG, Hatanaka E, Sato FT, Sampaio SC, Curi R. Suppressive effect of short-chain fatty acids on production of proinflammatory mediators by neutrophils. J Nutr Biochem 2011;22(9):849–55.

[105] Pouteau E, Meirim I, Métairon S, Fay LB. Acetate, propionate and butyrate in plasma: determination of the concentration and isotopic enrichment by gas chromatography/mass spectrometry with positive chemical ionization. J Mass Spectrom 2001;36(7):798–805.

[106] Rossi M, Amaretti A, Raimondi S. Folate production by probiotic bacteria. Nutrients 2011;3 (1):118–34.

[107] Strozzi GP, Mogna L. Quantification of folic acid in human feces after administration of bifidobacterium probiotic strains. J Clin Gastroenterol 2008;42(September):S179–84.

[108] Joubert BR, Den Dekker HT, Felix JF, Bohlin J, Ligthart S, Beckett E, et al. Maternal plasma folate impacts differential DNA methylation in an epigenome-wide meta-analysis of newborns. Nat Commun 2016;7(May 2015).

[109] Chang H, Zhang T, Zhang Z, Bao R, Fu C, Wang Z, et al. Tissue-specific distribution of aberrant DNA methylation associated with maternal low-folate status in human neural tube defects. J Nutr Biochem 2011;22(12):1172–7.

[110] Cho CE, Sánchez-Hernández D, Reza-López SA, Huot PSP, Kim YI, Anderson GH. High folate gestational and post-weaning diets alter hypothalamic feeding pathways by DNA methylation in Wistar rat offspring. Epigenetics 2013;8(7):710–9.

[111] Håberg SE, London SJ, Stigum H, Nafstad P, Nystad W. Folic acid supplements in pregnancy and early childhood respiratory health. Arch Dis Child 2009;94(3):180–4.

[112] Dunstan JA, West C, McCarthy S, Metcalfe J, Meldrum S, Oddy WH, et al. The relationship between maternal folate status in pregnancy, cord blood folate levels, and allergic outcomes in early childhood. Allergy 2012;67(1):50–7.

[113] Whitrow MJ, Moore VM, Rumbold AR, Davies MJ. Effect of supplemental folic acid in pregnancy on childhood asthma: a prospective birth cohort study. Am J Epidemiol 2009;170(12):1486–93.

[114] Amarasekera M, Martino D, Ashley S, Harb H, Kesper D, Strickland D, et al. Genome-wide DNA methylation profiling identifies a folate-sensitive region of differential methylation upstream of ZFP57-imprinting regulator in humans. FASEB J 2014;28(9):4068–76.

[115] Pompei A, Cordisco L, Amaretti A, Zanoni S, Raimondi S, Matteuzzi D, et al. Administration of folate-producing bifidobacteria enhances folate status in Wistar rats. J Nutr 2007;137(12):2742–6.

[116] Isolauri E. Role of intestinal flora in the development of allergy. Curr Opin Allergy Clin Immunol 2003;15–20.

[117] Blümer N, Sel S, Virna S, Patrascan CC, Zimmermann S, Herz U, et al. Perinatal maternal application of Lactobacillus rhamnosus GG suppresses allergic airway inflammation in mouse offspring. Clin Exp Allergy 2007;37(3):348–57.

[118] Braun-Fahrländer C, Gassner M, Grize L, Neu U, Sennhauser FH, Varonier HS, et al. Prevalence of hay fever and allergic sensitization in farmer's children and their peers living in the same rural community. Clin Exp Allergy 1999;29(1):28–34.

[119] Riedler J, Eder W, Oberfeld G, Schreuer M. Austrian children living on a farm have less hay fever, asthma and allergic sensitization. Clin Exp Allergy 2000;30(2):194–200.

[120] Von Ehrenstein OS, et al. Reduced risk of hay fever and asthma among children of farmers. Clin Exp Allergy 2000;30:187–93.

[121] Ernst P, Cormier Y. Relative scarcity of asthma and atopy among rural adolescents raised on a farm. Am J Respir Crit Care Med 2000;161(5):1563–6.

[122] Majkowska-Wojciechowska B, Pełka J, Korzon L, Kozłowska A, Kaczała M, Jarzębska M, et al. Prevalence of allergy, patterns of allergic sensitization and allergy risk factors in rural and urban children. Allergy 2007;62(9):1044–50.

[123] Korthals M, Ege MJ, Tebbe CC, von Mutius E, Bauer J. Application of PCR-SSCP for molecular epidemiological studies on the exposure of farm children to bacteria in environmental dust. J Microbiol Methods 2008;73(1):49–56.

[124] Debarry J, Hanuszkiewicz A, Stein K, Holst O, Heine H. The allergy-protective properties of *Acinetobacter lwoffii* F78 are imparted by its lipopolysaccharide. Allergy 2010;65(6):690–7.

[125] Brand S, Teich R, Dicke T, Harb H, Yildirim AÖ, Tost J, et al. Epigenetic regulation in murine offspring as a novel mechanism for transmaternal asthma protection induced by microbes. J Allergy Clin Immunol 2011;128(3).

[126] Suez J, Zmora N, Segal E, Elinav E. The pros, cons, and many unknowns of probiotics. Nat Med 2019;25.

[127] Liang L, Ai L, Qian J, Fang JY, Xu J. Long noncoding RNA expression profiles in gut tissues constitute molecular signatures that reflect the types of microbes. Sci Rep 2015;5(June):1–8.

This page intentionally left blank

CHAPTER 7

Sequencing technologies for epigenetics: From basics to applications

Rosario Michael Piro[1]
Department of Mathematics and Computer Science, Freie Universität Berlin, Berlin, Germany
Institute of Medical Genetics and Human Genetics, Charité-Universitätsmedizin Berlin, Berlin, Germany
German Cancer Consortium (DKTK) partner site Berlin, and German Cancer Research Center (DKFZ), Heidelberg, Germany

Contents

Introduction to high-throughput sequencing 163
    Next-generation sequencing 163
Library preparation 164
Flow cell preparation 165
Sequencing by synthesis 167
    Third-generation sequencing 168
Applications of sequencing technologies for epigenetics 169
    DNA methylation 170
    Histone modifications 171
    Other applications 171
Data processing and computational analysis 172
    Raw data and quality control 172
    Read alignment 173
    Analysis of methylation data 174
    Analysis of ChIP-seq data 176
    Analysis of data from other applications 177
Future perspectives 178
References 178

Epigenetic modifications of DNA and certain DNA-associated proteins can have a profound impact on gene regulation, the two most widely studied phenomena being DNA methylation and histone modifications [1–4]. Although strictly speaking epigenetic or epigenomic phenomena would be only those which confer sequence-independent, self-perpetuating, or heritable properties, such as stable DNA methylation [5–8], a much wider range of cellular marks with an impact on gene expression, including histone modifications, are often loosely referred to when using the terms "epigenetics" and "epigenomics" [8, 9]. Since ultimately, gene expression is influenced by a combination of multiple molecular modifications,

---

[1] Current address: Department of Electronics, Informatics and Bioengineering (DEIB), Polytechnic University of Milan, Milan, Italy.

Epigenetics of the Immune System
https://doi.org/10.1016/B978-0-12-817964-2.00007-1
© 2020 Elsevier Inc.
All rights reserved.

whether they are heritable or self-perpetuating in nature or not, in the following I will use this broader sense of the terms, well aware that they are only partially appropriate.

In mammalian cells, the predominant form of methylated DNA consists of 5-methylcytosine (5mC), i.e., a methyl group covalently bound to the fifth carbon of cytosine [10]. As this happens most frequently in the context of CpG dinucleotides (CpGs), which can be methylated on both complementary strands, the current model suggests that DNA methylation patterns are maintained during mitosis by methyltransferase DNMT1 which after DNA replication restores the complementary 5mCs of both hemi-methylated DNA copies [11]. While the function of DNA methylation is often oversimplified as a mark for transcriptional repression, it rather seems to be dependent on the genomic context in which it happens and may influence various biological processes such as transcriptional elongation, alternative splicing, alternative promoter usage, and the repression of repetitive DNA from intragenomic parasites [12].

DNA is wrapped around histone octamers—protein complexes composed of four types of subunits: histones 2A, 2B, 3, and 4 (abbreviated as H2A, H2B, H3, H4) [13]—and packaged in more or less densely compacted chromatin, a nucleoprotein complex structure whose architecture, and hence accessibility, is regulated by histone modifications [4]. Apart from their association with chromatin structure, some histone modifications at specific amino acid residues can either positively or negatively regulate the binding of transcription factors [4, 9, 14, 15], such that the combinatorial effect of specific histone modifications—described as the "histone code"—is thought to be consistent with determined transcriptional programs [16–18]. Well-known examples for histone modifications which are associated with transcriptional activity are H3K27ac (acetylation of lysine 27 of H3) as a mark of transcriptionally active regions, H3K27me3 (trimethylation of lysine 27 of H3) as mark of transcriptional repression, and H3K4me3 (trimethylation of lysine 4 of H3) as mark of active gene promoters. Apart from acetylation and methylation, histones can also be altered through phosphorylation, ubiquitination, and other chemical modifications.

DNA methylation, histone modifications, and other epigenetic mechanisms can be studied at genome scale using modern, massively parallel high-throughput sequencing technologies. Although these technologies have emerged less than two decades ago, they have already had a huge impact on many fields of biomedical research, including immunology [19], and are more and more being utilized also for clinical applications, e.g., for diagnostic purposes in cancer care. Despite their rather short history, various experimental approaches for high-throughput sequencing have been developed—all having their characteristic advantages and shortcomings—and due to a rapid succession of increasingly potent techniques, several of them have already fallen into disuse.

This chapter aims at providing an introductory, necessarily incomplete, overview of the high-throughput sequencing protocols, and computational methods most commonly used for characterizing the epigenetic regulation of cell samples in the context of immunological research. Most of the descriptions and considerations are, however, more generic and apply also to other research fields in which the epigenetic states of cells play an important role.

In the following sections, I shall first describe some notions of current next-generation and third-generation sequencing technologies ("Introduction to high-throughput sequencing" section), followed by their major applications in epigenetics research ("Applications of sequencing technologies for epigenetics" section). Apart from the technological aspects of sequencing, a second focus lies on computational methods for processing and analyzing sequencing data ("Data processing and computational analysis" section) before concluding the chapter with a brief outlook on future perspectives ("Future perspectives" section).

Introduction to high-throughput sequencing

The DNA sequencing method developed in the 1970s by Fred Sanger and his colleagues [20, 21] was one of the revolutionizing milestones of molecular biology. Improved and automated versions of Sanger sequencing are still frequently used, although their necessity for validating variants identified by next-generation sequencing has been questioned in the light of recent advances in these high-throughput technologies [22]. While Sanger sequencing has the advantage of being highly accurate, only the massively parallel "next-generation" sequencing technologies introduced less than two decades ago have made genome-wide analyses for large-scale projects possible. One of the major shortcomings of next-generation sequencing, namely its limitation of generating rather short sequence "reads" of at most a few hundred nucleotides length, has recently been addressed by the introduction of so-called "third-generation" sequencing methodologies.

For a proper computational analysis and meaningful interpretation of sequencing data, it is essential to understand how the data to be analyzed have been generated. Only then can intrinsic sources of errors, artifacts, limitations, and biases be efficiently addressed. An exhaustive discussion of current and emerging sequencing technologies, however, would easily exceed the scope of this chapter. Therefore, in the following, I will mainly concentrate on the major next-generation sequencing platform because it is widely used in epigenetic studies [9]. Nonetheless, also third-generation sequencing and its potential impact shall be briefly introduced.

Next-generation sequencing

High-throughput next-generation sequencing provides a powerful tool for genome-wide analyses of molecular components at different levels of cellular organization and function, e.g., genomic DNA sequences, RNA sequences, and transcript levels, the methylome, binding sites of DNA-associated proteins, and three-dimensional chromosome architecture.

The revolutionizing innovation introduced by next-generation sequencing techniques was the simultaneous sequencing of millions or even billions of short DNA fragments in parallel. While each individual sequence "read" (about 25–400 bp) generated has a much lower accuracy than Sanger sequencing, in principle the sheer amount of sequenced DNA fragments allows for an accurate determination of nucleotide sequences because in most cases each individual base is sequenced by multiple reads such that sequencing errors on individual reads have little negative impact on the determined consensus sequence.

Several different experimental protocols and methods for high-throughput sequencing have been developed in the past, such as Life Technology's SOLiD system, Roche's 454 sequencing system, and the Illumina Genome Analyzer, HiSeq, and NovaSeq platforms. Many of the older sequencing platforms, however, are no longer produced or rarely utilized nowadays. For a comprehensive summary and comparison of some of these platforms, please see Metzker's review article of 2010 [23]. In the following, I will describe Illumina sequencing, which is by far the most commonly used technology for the majority of next-generation sequencing applications.

Illumina sequencing protocol

Many of the experimental steps of a sequencing protocol can lead to characteristic errors in the sequencing data. Therefore, it is always advisable to familiarize with a sequencing protocol before processing the produced data. The basic Illumina sequencing protocol used to sequence DNA fragments has three major steps that I shall summarize here (for more details, see for example Robinson et al. [24]).

Library preparation

The preparation of the sequencing library is illustrated in Fig. 1. For most next-generation sequencing applications, DNA molecules are first sheared into random fragments by sonication, which utilizes ultrasound waves in a solution to break DNA, or by nebulization, which repeatedly forces the DNA solution as a fine mist through a small hole. For some applications, however, enzymatic digestion at precise sequence patterns is preferred to a random fragmentation, e.g., for the chromosome conformation capture protocol Hi-C [25].

After fragmentation, single-stranded random overhangs at the ends of the DNA fragments are repaired ("end repair") in order to have blunt ends, and a single adenine ("A overhang") is added to both 3′ ends (Fig. 1B). This prevents DNA fragments from ligating to one another and allows for ligation of sequencing adapters ("adapter ligation"), which are designed to have a complementary thymine (T) base overhang, to both ends of each fragment (Fig. 1C).

The chemically synthesized Illumina sequencing adapters provide specific sequences that allow for PCR enrichment of adapter-ligated DNA fragments, binding to complementary oligonucleotides on the sequencing flow cell (see below), and optional "barcoding" for simultaneous sequencing of multiple samples. Since there are different

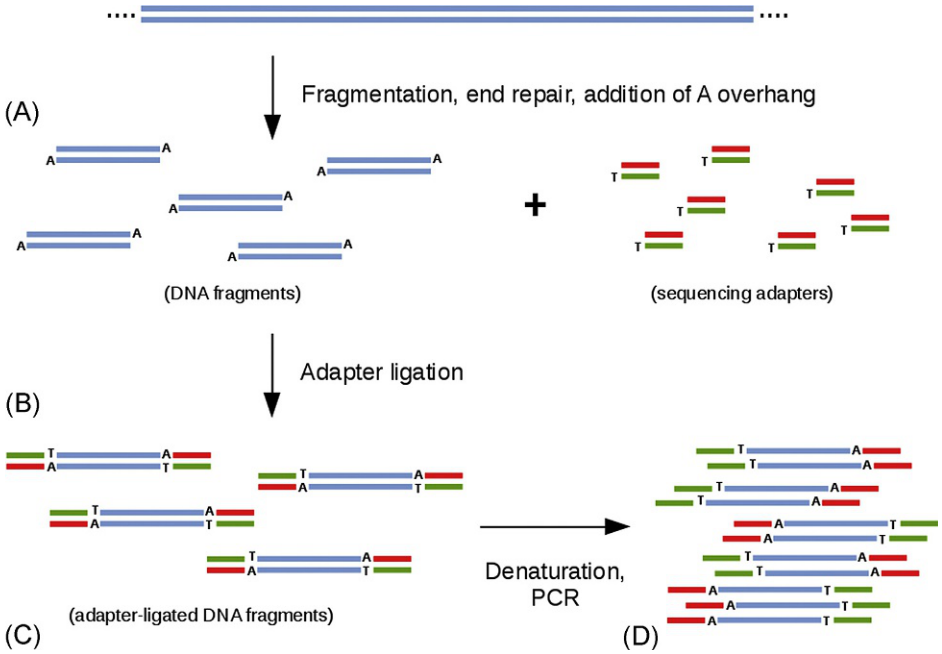


Fig. 1 Library preparation for Illumina sequencing (simplified). Input DNA (A) is fragmented, ends are repaired and a single adenine is added to the 3′ of both strands (B). Sequencing adapters are added and ligated to the DNA fragments (C). Double-stranded, adapter-ligated fragments are denaturated and enriched by PCR (D).

Illumina adapter sequences for different purposes (e.g., single-end sequencing, paired-end sequencing, etc.), users should consult the producer's documentation for details.

After the ligation reaction, purification and size selection help to remove unligated adapters and adapter dimers (adapters that have ligated to one another) and to obtain a library of fragments of comparable sizes. The final library preparation step consists of a few cycles of enrichment PCR using primers which anneal to the adapter ends. This selectively enriches those DNA fragments that have ligated adapter molecules on both ends and increases the chances for a particular sequence fragment to attach to the flow cell in the following major step of the sequencing protocol. At the same time, the enrichment PCR is the main cause of the so-called "duplicate reads," i.e., multiple reads which stem from exactly the same original DNA fragment and therefore provide redundant information that has to be handled accordingly in downstream analyses.

Flow cell preparation

An Illumina flow cell is a hollow glass slide with one or more channels ("lanes"), coated with oligonucleotides which are complementary to the sequencing adapters so that single-stranded, adapter-ligated DNA fragments can attach through hybridization (see Fig. 2). While the first Illumina sequencing platforms worked with randomly coated flow cells, recent enhancements (Illumina HiSeq X Ten, HiSeq 3000/4000, and NovaSeq) have introduced patterned flow cells with ordered wells designed for optimal spacing and uniform sizes of spots for DNA fragment deposition.

After preparation of the fragment library as described above, the double-stranded fragments are denaturated with NaOH and the single-stranded molecules are deposited on the flow cell where they randomly attach to the complementary oligonucleotides. Since they are not covalently fixed to the flow cell and might thus be washed away, their

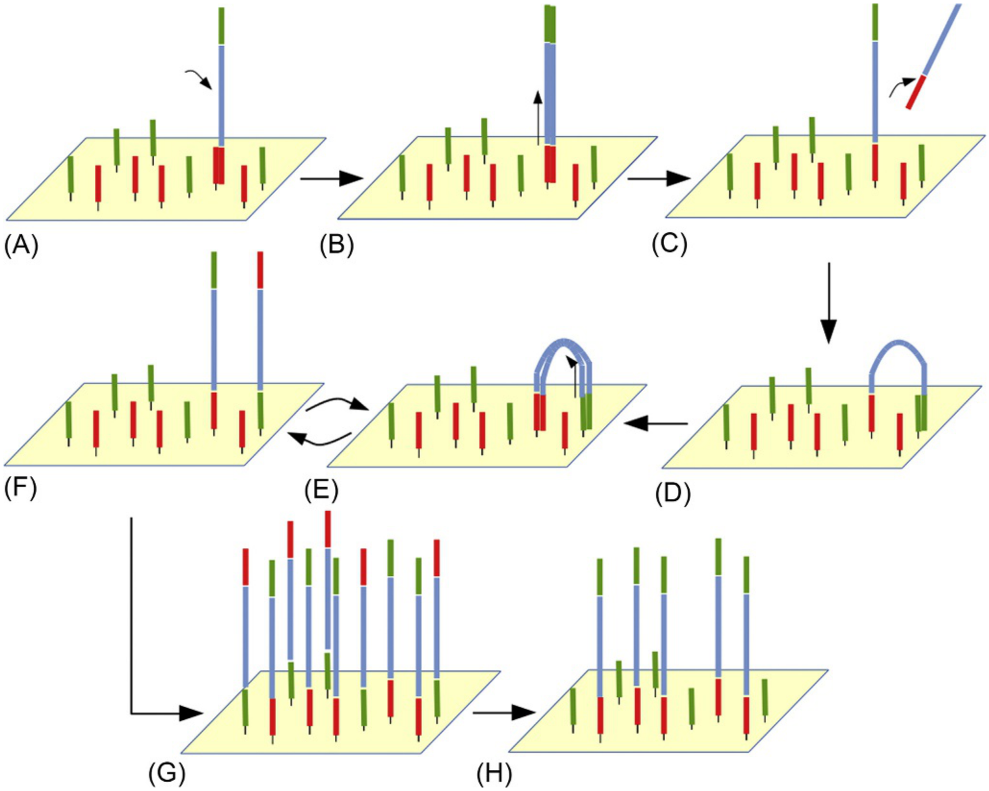


Fig. 2 Flow cell preparation for Illumina sequencing (simplified). DNA fragments are hybridized to oligonucleotides on the flow cell (A). Covalently fixed complementary strands are synthesized starting from the oligonucleotides (B) and the original fragments are washed away (C). For bridge amplification, DNA fragments are bent by hybridizing their second adapter to another oligonucleotide on the flow cell (D), subsequently, their complementary strands are synthesized beginning from the second oligonucleotide (E), and finally, the two strands are separated by denaturation (F). This is repeated to form a cluster of DNA fragments (G) of which one of the two orientations (complementary strands) is finally cleaved (H). In this figure, 5′ and 3′ sequencing adapters and their complementary oligonucleotides on the flow cell are depicted in the same color.

complementary sequences are constructed by elongating the fixed, complementary oligonucleotides to which they had hybridized. Then, the obtained double-stranded DNA products are denatured and the original single-stranded fragments are washed away; only their covalently bound complements remain (Fig. 2A–C).

Individual fragments, however, would not yield a signal strong enough to be detected. Hence, "bridge amplification" is used to create small clusters, or colonies, of identical fragments by repeatedly bending each fragment, such that also its second adapter hybridizes to an oligonucleotide in its vicinity, and using it as a template to create the complementary sequence by extending the second oligonucleotide (Fig. 2D–F). This form of PCR, however, creates colonies with amplicons in both orientations (i.e., $5'$ to $3'$ and $3'$ to $5'$) of the original single-stranded DNA fragment. Therefore, one of the two strands is cleaved and discarded before sequencing (Fig. 2G and H), leaving a cluster of identical, single-stranded DNA templates for each DNA fragment which originally hybridized to the flow cell.

Sequencing by synthesis

The sequencing of the identical DNA fragments of each cluster is then performed by the simultaneous synthesis of the complementary strands of the single-stranded fragments (see Fig. 3). The primers required for starting the synthesis of the complementary strand correspond to a specific region of the sequencing adapter which was ligated to every original DNA fragment.

In every sequencing cycle a single, fluorescent dye-labeled nucleotide is added to each growing complementary strand (Fig. 3B and C), and upon laser excitation, the emitted, base-specific wavelength of the added nucleotide is registered (Fig. 3D). To ensure that only one base is added at a time, the procedure relies on deoxy-nucleoside triphosphate (dNTP) terminators, i.e., nucleotides with a reversible block of the $3'$-OH group, such that a further extension of the growing nucleotide chain in the following cycle is possible only after removal of the fluorescent block (Fig. 3E and F). Recent Illumina platforms (NextSeq and NovaSeq) use two instead of four channels (relying on red and green filters) for sequencing by synthesis such that two images are sufficient to determine all four base calls. This speeds up the entire procedure.

Since the quality of the fluorescence signals depends on various factors—such as PCR errors during bridge amplification or phase errors due to nonincorporated bases during sequencing—for each base call, a quality score is determined to measure the likelihood of an erroneous base call.

When performing paired-end sequencing (one read for each end of a DNA fragment), the first sequencing product is washed away after denaturation. Then, the reverse complement of the entire DNA fragment is generated and can subsequently be sequenced in the same way to obtain the second read from the other end of the DNA fragment.

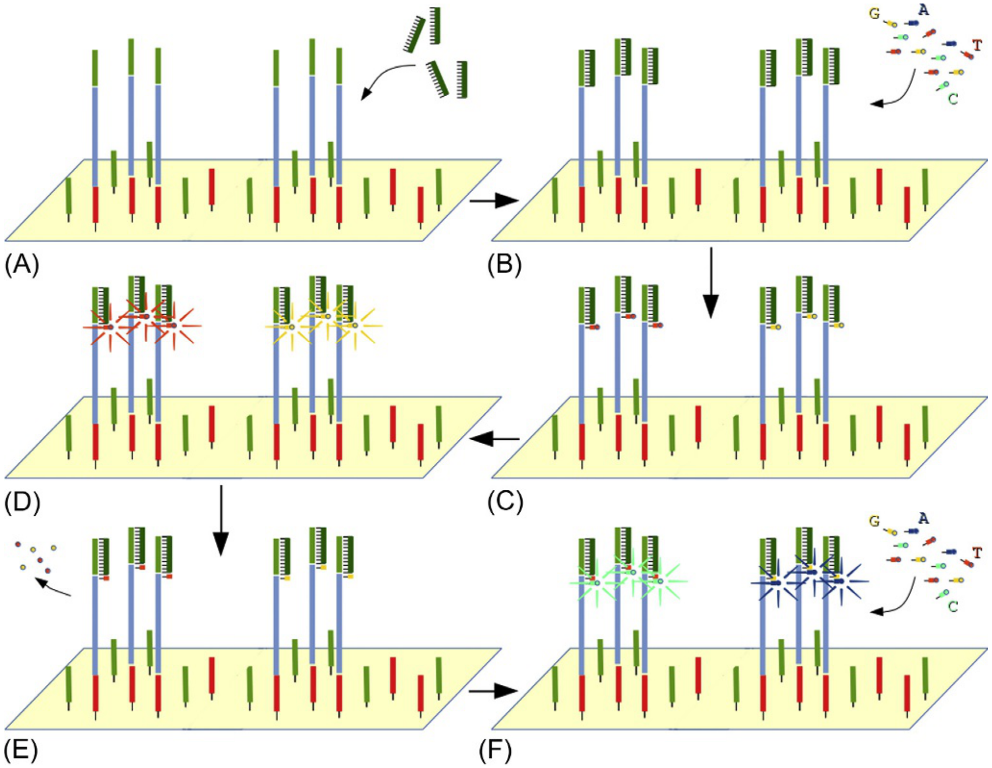


Fig. 3 Sequencing by synthesis (simplified). Two fragment clusters are shown. First, sequencing primers are hybridized to the fragments' sequencing adapters (A). In each sequencing cycle, labeled dNTP terminators of all four bases are added (B) and are used to extend the complementary sequence of the DNA fragments (C). The fluorescent labels of the inserted bases are registered upon laser excitation (one image per channel) (D) and the blocks are removed (E) such that the complementary strands can be extended by another base in the next cycle (F).

Third-generation sequencing

One of the major limitations of next-generation sequencing is the rather short read length of at most a few hundred nucleotides. This can be problematic, for example, for the identification of certain structural variants in whole-genome sequencing data. Also, the inference of complete transcript isoforms from short-read RNA-seq data [26] poses a considerable problem [27]. For some applications, it would, therefore, be advantageous to be able to sequence longer reads.

Another disadvantage of next-generation sequencing is its reliance on extensive library preparation steps, including, for example, PCR amplification which may lead to undesired effects like duplicate reads, nonlinear amplification of different transcript levels in RNA-seq experiments, and other biases.

Recently developed "third-generation" or "long-read" sequencing technologies [28] address these issues by implementing direct single-molecule sequencing with substantially longer reads than next-generation sequencing methods. Despite some disadvantages with respect to next-generation sequencings, such as higher error rates [29], some of these novel sequencing methods have already been adopted in epigenetics research [30]. Here, I briefly describe two prominent examples: nanopore sequencing and single-molecule, real-time (SMRT) sequencing.

Nanopore sequencing

Oxford Nanopore Technologies (ONT) was the first to introduce a commercial nanopore sequencer in 2014 [31]. The concept underlying this technology is the observation of changes in ion currents when the nucleotides of a single-stranded DNA molecule are pulled through a tiny bacterial pore ("nanopore") by a phage DNA polymerase [32, 33].

Since epigenetically modified DNA bases, such as 5mC, lead to differences in currents, they can be distinguished from their unmodified counterparts [34, 35].

SMRT sequencing

Pacific BioSciences (PacBio) developed an approach whose purpose is to determine sequence information during the replication of a long, circular, and single-stranded DNA template molecule. This DNA template is produced from a double-stranded DNA molecule after ligation of hairpin adapters to both ends. Sequencing is then performed by adding the four nucleotides labeled with different fluorescent dyes and monitoring the activity of DNA polymerase during the replication of the template's complementary strand. When a base is loaded by the DNA polymerase a light pulse is emitted [29].

As with nanopore sequencing, an advantage of this approach is that epigenetic base modifications such as DNA methylation change the kinetics of, i.e., the time required for, the incorporation of a base and can thus be determined along with the plain sequence information [36].

Applications of sequencing technologies for epigenetics

How exactly sequencing technologies can be applied for studying epigenetic marks in immunological research and other fields depends very much on the different technologies themselves. While, for example, as mentioned above third-generation technologies can directly determine some DNA base modifications, next-generation sequencing usually requires additional experimental, biochemical steps to prepare samples for the specific sequencing purpose. Here, I shall concentrate mostly on the application of next-generation sequencing because to date these technologies are much more frequently used for epigenetics research.

DNA methylation

Many different sequencing protocols have been developed to determine nucleotides within a DNA molecule which have been epigenetically modified by covalently adding methyl groups [9, 30, 37]. The most prominent and widely used methodology for obtaining methylation profiles at single base-pair resolution is whole-genome bisulfite sequencing (WGBS or methylC-seq) [38]. WGBS relies on converting cytosines into thymines (via uracil) by the treatment of genomic DNA with sodium bisulfite. Methylcytosines instead are protected from conversion [39] and can thus be easily identified after sequencing. For cases where most regions of high CpG content (such as promoters) are of interest, rather than CpG-deficient intergenic and intronic regions, reduced representation bisulfite sequencing (RRBS) [40] allows to significantly lower the sequencing effort and cost by combining the digestion of genomic DNA by restriction enzymes, subsequent fragment size selection, and finally bisulfite sequencing.

Since these methods, however, require the conversion of unmethylated cytosines to uracil as an essential library preparation step, methylation levels are often overestimated in case of incomplete conversion [41, 42]. If an unmethylated cytosine is not converted to uracil, it will be wrongly interpreted as a methylated cytosine (5mC). This can be caused, for example, by incomplete denaturation of the double-stranded DNA because only the cytosines of single-stranded molecules are converted upon bisulfite treatment [43].

Another confounding factor is the presence of 5-hydroxymethylcytosine (5hmC), an intermediate formed during DNA demethylation, which is converted to cytosine-5-methylsulfonate by bisulfite and subsequently read as a regular cytosine during sequencing. Converted 5hmC is hence indistinguishable from nonconverted 5mC [44]. This problem is addressed by oxidative bisulfite sequencing (oxBS-seq) [45], which relies on chemical oxidation of 5hmC to 5-formylcytosine (5fC) that, like unmethylated cytosine, can subsequently be converted to uracil by bisulfite sequencing. Thus, in principle, only 5mC is protected and read as cytosine during sequencing. Optionally, levels of 5hmC can be quantified by performing both WGBS and oxBS-seq and measuring the difference in observed methylation levels. Apart from oxBS-seq, also other protocols, such as Tet-assisted bisulfite sequencing (TAB-seq) [46, 47], have been established to be able to discriminate between 5mC and 5hmC.

As illustrated by these brief summaries, the most prominent library preparation procedures of next-generation sequencing methods for DNA methylation profiling tend to require an additional chemical treatment of the sample which is critical for the quality and correct interpretability of the obtained results. Third-generation sequencing (see "Third-generation sequencing" section) on the other hand provides an opportunity for methylation analysis directly from DNA molecules without the need for previous chemical conversion [30, 34–37]. These technologies are still less accurate and less mature than next-generation sequencing, but being rapidly improved they will likely become more and more attractive for DNA methylome analysis.

Histone modifications

Patterns of histone modification [4] can be interrogated with the same genome-wide techniques that are used to identify the physical protein-DNA interactions of a specific protein of interest, such as the binding sites of a specific transcription factor [48]. The standard methodology for this purpose is ChIP-seq which combines chromatin immunoprecipitation (ChIP) with high-throughput DNA sequencing [48].

To prepare a sequencing library for a ChIP-seq experiment, first, all DNA-associated proteins are covalently cross-linked to the DNA they have bound, typically by treating cells with formaldehyde before DNA extraction. Then, DNA is sheared into fragments as in a typical next-generation sequencing experiment and subsequently, an antibody specific to the protein of interest is used to selectively co-immunoprecipitate only those protein-bound DNA fragments which are cross-linked to the specific protein of interest. Finally, the protein-DNA links of the immunoprecipitated fragments are reversed, the proteins are washed away, and the purified DNA fragments are prepared for sequencing in order to determine to which genomic loci the protein of interest was bound. This procedure can easily be adapted to histone modification analysis by utilizing specifically designed antibodies that bind only to histones which carry specific modifications [48].

The sensitivity of ChIP-seq is strongly related to the number of generated sequence reads whose origins can be pinpointed in the reference genome, i.e., to the depth of the sequencing run. Inferred binding sites often diverge only a few tens of nucleotides from the actual binding sites of the protein of interest. Such a rather high accurate localization can be achieved by integrating information from a large number of individual sequence reads which cluster around the true binding sites [49]. The resolution can further be improved using chromatin immunoprecipitation-exonuclease (ChIP-exo) where the ends of immunoprecipitated DNA fragments are first shortened by exonuclease digestion such that the obtained reads are closer to the protein's true binding site [50].

ChIP-seq experiments are usually performed along with an additional control experiment which is performed without the specific antibody. This helps to estimate and eliminate bias due to GC content, read mappability, DNA repeats, and copy number variants present in the genome of interest.

Other applications

A wide range of other sequencing protocols and methods exists for the genome-wide investigation of epigenetic and other biological processes [9]. Apart from those described above, some frequently used techniques are

- RNA-seq [26] to analyze gene activity at the transcriptomic level;
- Hi-C [25], ChIA-PET (chromatin interaction analysis by paired-end tag sequencing) [51] and other chromatin conformation capture methods to study the 3D chromatin structure, i.e., the physical short- and long-range interactions between different genomic regions;

- sequencing after micrococcal nuclease (MNase) digestion of all nonbound linker DNA between nucleosomes (MNase-seq) [52] to infer nucleosome positioning and occupancy; and
- DNase hypersensitivity analysis by sequencing after DNase I digestion of free, unwound DNA in nucleosome-depleted regions (DNase-seq) [53, 54] for measuring chromatin accessibility.

These and other techniques measure either directly gene transcription levels or interrogate molecular marks which can in one way or another influence transcription rates of genes and are therefore considered as epigenetic marks [9] ("epigenetic" in a broad sense, as defined at the beginning of this chapter, although they are not necessarily self-perpetuating or heritable). Of course, sequencing may be useful for studying also other aspects of the immune system, apart from epigenetic/epigenomic mechanisms, e.g., for determining the immunological repertoire and for antibody discovery and engineering [19, 55, 56].

For some of these applications, third-generation sequencing might soon become an important alternative to currently used next-generation techniques. RNA-seq, for example, as already mentioned suffers from the difficulty of accurately reconstructing complete transcript isoforms due to the short-read length [27], a problem that can be better addressed using long-read sequencing [57].

Data processing and computational analysis

Appropriate computational processing and analysis is fundamental for a correct interpretation of the raw data generated by next-generation and third-generation sequencing [42, 58, 59]. Two aspects are particularly important: (i) data quality, the probably most important law in bioinformatics being "garbage in, garbage out" and (ii) the choice of the right analysis approach because the sequencing "reads" produced by different techniques and protocols may have exactly the same format but nonetheless largely different meaning. The following discussion is intended to give a few pointers without any claim of completeness.

Raw data and quality control

Next-generation sequencing has exponentially increased the throughput and thus the amount of sequence information available, and at the same time, the sequencing cost has massively dropped. This, however, comes at the expense of accuracy with respect to much more precise low-throughput methodologies like Sanger sequencing. The quality of next-generation sequencing data can be affected by many error sources such as errors during PCR or bridge amplification, or decaying base call qualities due to so-called "phase errors," i.e., increasingly inconsistent light signals from a fragment colony or cluster on the flow cell because individual molecules did not incorporate the nucleotide at some previous sequencing cycle [19]. Miscalled bases could, for example, erroneously be counted as cytosine to thymine conversions when assessing methylation status from bisulfite sequencing data. Therefore, the quality of the initial raw sequence data (but also of the intermediate and final results of the analysis workflow) should be thoroughly controlled [42, 60].

Raw sequencing data from Illumina sequencers are typically stored in FASTQ format [61] where each read, that is, the sequence of bases that have been called from the light signals of an individual fragment colony, is associated with a sequence of corresponding base quality scores of equal length and a unique read identifier. Base quality scores are encoded as ASCII characters and represent the probability of an erroneous base call, i.e., the confidence level of the base [24, 58].

The foremost (but not the only) goal of quality control of raw sequence data is to verify the base and read qualities prior to preprocessing and computational analysis [24, 58]. Overall read quality can be improved by filtering low-quality reads, removal of contaminating adapter sequences, and sometimes trimming of low-quality bases from individual reads [24, 58].

Many QC tools, like FastQC [62], BIGpre [63], and PIQA [64], can be used for a wide range of next-generation sequencing applications but would erroneously consider the significant lack or abundance of thymine and cytosine bases (depending on methylation status) typical of bisulfite sequencing as bias [58]. Examples of QC tools that can handle such data and assess the efficiency of bisulfite treatment are MethyQA [65] and BSeQC [66].

Some of these quality control tools, such as FastQC, have been adapted to be able to handle long reads from third-generation sequencing, too, and other dedicated tools have also recently been developed for this purpose [67, 68].

Read alignment

Once raw data quality has been verified, reads are usually aligned, or "mapped," to the reference genome to identify their chromosomal location of origin. In this step, it is fundamental to use an appropriate read aligner, according to the experimental protocol that has been used for sequencing.

For many next-generation sequencing applications, such as standard DNA sequencing, ChIP-seq, and Hi-C, conventional alignment tools like BWA [69, 70], BWA-MEM [71], Bowtie 2 [72], or SOAP2 [73] are suitable. Two important exceptions are transcriptomic RNA-seq data because many reads contain exon-exon junctions which can translate to possibly very large gaps (introns) when aligned to the reference genome [74], and data from bisulfite sequencing, because of the high dissimilarity between the reference genome and sequences obtained from bisulfite-treated DNA fragments [42, 58, 59, 75].

Not all conventional aligners can handle these particular cases, but several dedicated aligners have been published, such as STAR [76] which is a good example for a prominent RNA-seq read aligner.

Some bisulfite sequencing read aligners like RMAP-bs [77], BSMAP [78], and Segemehl [79] substitute thymines in the reads for "wild-card" Y nucleotides (which conventionally stand for either cytosine or thymine) prior to read mapping because each thymine could potentially stem from a converted unmethylated cytosine. An alternative approach, taken by aligners like MethylCoder [80], BS-Seeker [81], BS-Seeker2 [82], BRAT (Bisulfite-treated Reads Analysis Tool) [83], BRAT-BW (BRAT using the Burrows-Wheeler transform) [84], bwa-meth [85], and Bismark [86], is to convert cytosines to thymines in both the reads and the reference genome, basically performing read mapping with a reduced alphabet of only three nucleotides. The performance and accuracy of many of these bisulfite sequencing aligners are comparable and none of them seems significantly better or worse than competing tools [42].

Also, the alignment of long reads from third-generation sequencing may require appropriate computational approaches. A recent example of a dedicated long-read aligner is Minimap2 [87].

Read alignments are commonly specified and stored in the Sequence Alignment/Map (SAM) format [88, 89] or its compressed, binary version Binary Alignment/Map (BAM). Both can be sorted according to chromosomal location and indexed for viewing and quick and efficient searches. The SAM format allows to specify a wide range of useful, alignment-related information, including, for example, the chromosomal locations to which reads have been aligned, their mapping quality, which read bases are aligned to the reference genome and where gaps in the alignment (possible indels) are located, and whether a read maps to the forward or the reverse strand of the reference genome. For efficiently working with sequencing data and better understanding the rationale behind many downstream tools, it is advisable to familiarize with the SAM/BAM format [24, 89].

In many cases read alignments will require some postprocessing, e.g., to sort and index the aligned reads, to mark or remove duplicate reads, or to recalibrate base quality scores. For further details, please see Robinson et al. [24] and Van der Auwera et al. [90] or dedicated online sources.

Analysis of methylation data

Like for most sequencing applications and data processing steps, many different algorithms and tools for methylation profile segmentation and detecting differentially methylated cytosines (DMCs) and differentially methylated regions (DMRs) have been developed [42, 58]. As most approaches rely on bisulfite sequencing, the following descriptions and considerations are always referred to this methodology, except where explicitly stated otherwise.

DNA methylation scoring

Aligned reads have to be aggregated to provide a meaningful, interpretable assessment of DNA methylation states both of individual CpGs (or other contexts for some nonhuman data) as well as entire genomic regions [58].

The methylation status of an individual cytosine is frequently measured by the so-called β-score [91] which is simply defined as the proportion of reads, aligned to the corresponding genomic location, which after bisulfite treatment still contain a cytosine (methylated) at that position instead of thymine (unmethylated). Thus β-scores can range from 0 to 1 and reflect the fraction of cells in which the analyzed cytosines are methylated. A high accuracy, however, is possible only if the coverage (number of mapping reads) is sufficient, hence cytosines with poor depth of coverage are usually excluded from further analyses [58].

Several tools, including Bismark, BS-Seeker2, and BRAT-BW, provide β-scores for individual cytosines, along with information about the depth of coverage which is necessary to compute the statistical relevance of an observed methylation score. Methylation scores from individual cytosines can then be aggregated for regional DNA methylation scoring by binning the genome into windows and aggregating the β-scores within each bin [58], e.g., using MethylKit [92].

For an appropriate interpretation of the results, it is advisable to annotate methylation patterns with their genomic context (e.g., promoters, transcription start site regions, gene bodies, etc.), using tools like GBSA (Genome Bisulfite Sequencing Analyzer) [93] and WBSA (Web service for Bisulfite Sequencing data Analysis) [94]. For more details and further considerations about DNA methylation scoring and assessing DNA methylation according to the genomic context, please see Adusumalli et al. [58].

Lately, also tools to determine methylation status from third-generation sequencing data are being developed. AgIn [95], for example, has been designed to infer DNA methylation of CpGs from PacBio SMRT sequencing data, particularly keeping repetitive regions in mind which are notoriously difficult to map with short reads from next-generation sequencing.

Differential methylation

DNA methylation is known to influence transcriptional regulation of the immune system [96], and differential methylation has been shown to play important roles in human disease. Differential DNA methylation of several genes associated with inflammation, for example, may be linked to posttraumatic stress disorder [97]; poor-prognosis hindbrain ependymomas have been shown to often exhibit a CpG island methylator phenotype [98]; and a methylation signature of tumor-infiltrating lymphocytes has recently been proposed as a biomarker for survival and therapy response prediction in breast cancer [99].

Hence, a common purpose of many studies is the comparison of methylation patterns between different conditions [42, 58]. Given the particular nature of bisulfite sequencing data and the different levels of organization at which methylation patterns can be compared (individual cytosines, genomic regions of various sizes, etc.), it is often advisable to assess differential methylation with help of the tools that have been specifically developed for the purpose [58]. WBSA, for example, can identify DMRs but does not determine individual DMCs. MethylKit, instead, works for DMCs (or scalable bins) and MethPipe [100] can be used for both. Apart from these, many other tools have been developed for DMC and/or DMR detection; more information and suggestions can be found, for example, in review articles by Adusumalli et al. [58] and Wreczycka et al. [42].

Methylome segmentation

Methylation profiles and dynamics can not only be analyzed across samples to detect differential methylation but also within one sample, e.g., for the identification of unmethylated regions (UMRs), fully methylated regions (FMRs), low-methylated regions (LMRs), and sometimes partially methylated domains (PMDs) [42]. To this end, the methylome of an individual sample can be segmented into regions of similar methylation levels, e.g., using MethylKit, MethPipe, the R package MethylSeekR [101], or the approach of Stadler and colleagues [102].

Analysis of ChIP-seq data

Considering the experimental ChIP-seq procedure as described above, the net effect after having aligned the sequenced reads to the reference genome should be an accumulation, i.e., enrichment, of reads in "peaks" around the genomic locations to which the protein of interest had interacted with DNA. The computational identification of these locations based on the mapped reads is therefore usually referred to as "peak calling" [103, 104]. In case of antibodies targeting specific histone modifications, these peaks would indicate at which locations in the genome, the histone of interest carried said modification.

Generally, peak calling can be divided into two subproblems, the identification of peaks and the test of their significance [104]. Popular tools for peak calling from ChIP-seq data are MACS and MACS2 (Model-based Analysis of ChIP-Seq data) [105, 106], Dfilter [107], BCP (Bayesian Change-Point) [108], and MUSIC (MUltiScale enrIchment Calling for ChIP-Seq) [109]. Many other tools for peak calling or differential ChIP-seq analysis have been published since the first applications in 2007 [104, 110]. The main challenge of these methods is to reliably identify true peaks among the noise, i.e., among randomly distributed reads, with a reasonable trade-off between sensitivity (few false negatives) and specificity (few false positives). The choice of the method may depend on the actual type of ChIP-seq data to be analyzed. In a recent comparison of several tools, for example, BCP and MUSIC have performed best for histone data, and BCP and MACS2 for transcription factor binding data [104].

Quality control of ChIP-seq data

As for all sequencing applications, quality control of ChIP-seq data is essential to ensure a meaningful interpretability [111]. Several different QC measures are often employed to this end. Here, I shall briefly describe two of them for illustrative purposes. For an introduction to more sophisticated quality measures, please see Landt et al. [111]

One critical issue is library complexity. An elevated number of duplicate reads can indicate, for example, problems during the extraction of DNA from the cells or during library generation. One possible measure to analyze library complexity is the so-called nonredundant fraction (NRF) of uniquely mapped reads, that is, the proportion of these reads that have unique start positions in the genome. NRF decreases with sequencing depth because deeper sequencing increases the chance of two individual reads having the same start position without actually being PCR duplicates. The ENCODE project recommends a target NRF of ≥ 0.8 for 10 million uniquely mapped reads [111].

Another issue is the global enrichment of protein-bound DNA achieved by the ChIP experiment. Ideally, all sequenced DNA fragments should stem from regions to which the protein of interest did bind, but this is usually far from the truth. Indeed, typically a minority of reads in ChIP-seq experiments occur in significantly enriched genomic regions (i.e., in peaks); the remaining reads represent background noise. As a simple makeshift measure for the global enrichment of the immunoprecipitation, one can calculate the fraction of reads falling in peaks, FRiP, which has a rough correlation with the number of called peaks and for most ENCODE datasets indicate an enrichment of 1% or more when peaks are called using MACS with default parameters [111]. For histone modification data, however, where broader regions rather than individual peaks may show the enrichment of sequencing reads, more sophisticated approaches to model immunoprecipitation efficiency may sometimes be advantageous [112].

Analysis of data from other applications

Although bisulfite sequencing and ChIP-seq may be the major sequencing techniques for epigenetics, many other techniques and protocols are used in the context of, or in conjunction with, the identification of epigenetic marks. Moreover, sequencing techniques may of course also aid in immune system-related projects in which epigenetics plays no role. An exhaustive discussion of the analysis of data generated by these methods is prohibitive in this chapter, but at least a few considerations and examples shall be given here.

Apart from epigenetics, next-generation sequencing may help, for example, to analyze phage- and other antibody-display libraries [56], to identify functionally important genetic variation between individuals which is associated with immune function in health and disease [113], or to dissect tumor-immune cell interactions [114].

Also, transcriptomic studies may be of interest in immunological research. RNA-seq is most frequently used to quantify gene expression levels [19], e.g., for identifying differentially expressed genes using tools like DESeq2 [115], but the technique is actually much more powerful because it can provide information on alternative splicing (e.g., through splice junctions inferred from junction-spanning reads) [116, 117], intron retention (e.g., through reads which contain intronic sequences) [118], allelic imbalance (e.g., through reads containing heterozygous germline variants or somatic mutations) [119, 120], and so forth.

Also, chromosome conformation capture techniques such as Hi-C can be used, for example, to study how enhancers control the expression of immunity genes during virus infection [121] or to identify autoimmune disease-associated variants in CD4$^+$ T cells [122].

Future perspectives

As can be suspected from the discussion above, next-generation sequencing and its applications are still the gold standards when it comes to large-scale epigenetic research regarding the immune system. Given the very fast pace at which sequencing technologies have been continually improved, however, we should not be surprised if the accuracy and quality of third-generation sequencing data will see significant enhancements in the near future. It is therefore likely that third-generation sequencing techniques will find more and more employment in both basic research and practical applications.

Whatever sequencing technology is used, the extraordinary possibilities that high-throughput sequencing provides will hopefully help to gain a better understanding of the causal relationships between genetics on one hand and epigenetic events and other dynamic processes on the other, including DNA methylation, histone modifications, transcription factor binding, gene expression, and how genotypes can affect these [9, 123].

One major challenge for many of the techniques described above is the heterogeneity of a cell population from which sequencing data is usually obtained, which obscures both the interpretability and the generalizability of the results [9]. For this reason, single-cell epigenomic methodologies [124–126] are likely to gain in popularity as sequencing costs drop. Ideally, sequencing of multiple molecular types (e.g., methylated DNA, histone marks, RNA, and chromosome interactions) from the same single cell in parallel should be the final goal, which might benefit from current and upcoming third-generation sequencing technologies which can detect multiple chemical species at once, e.g., DNA methylation along with the DNA sequence [127].

References

[1] Jaenisch R, Bird A. Epigenetic regulation of gene expression: how the genome integrates intrinsic and environmental signals. Nat Genet 2003;33(Suppl):245–54.

[2] Gibney ER, Nolan CM. Epigenetics and gene expression. Heredity 2010;105(1):4–13.

[3] Cholewa-Waclaw J, Bird A, von Schimmelmann M, Schaefer A, Yu H, Song H, et al. The role of epigenetic mechanisms in the regulation of gene expression in the nervous system. J Neurosci 2016;36(45):11427–34.

[4] Bannister AJ, Kouzarides T. Regulation of chromatin by histone modifications. Cell Res 2011;21(3):381–95.

[5] Wolffe AP, Jones PL, Wade PA. DNA demethylation. Proc Natl Acad Sci U S A 1999;96(11):5894–6.

[6] Vandiver AR, Idrizi A, Rizzardi L, Feinberg AP, Hansen KD. DNA methylation is stable during replication and cell cycle arrest. Sci Rep 2015;5:17911.

[7] Hofmeister BT, Lee K, Rohr NA, Hall DW, Schmitz RJ. Stable inheritance of DNA methylation allows creation of epigenotype maps and the study of epiallele inheritance patterns in the absence of genetic variation. Genome Biol 2017;18(1):155.

[8] Ptashne M. Epigenetics: core misconcept. Proc Natl Acad Sci U S A 2013;110(18):7101–3.

[9] Sarda S, Hannenhalli S. Next-generation sequencing and epigenomics research: a hammer in search of nails. Genomics Inform 2014;12(1):2–11.

[10] Smith ZD, Meissner A. DNA methylation: roles in mammalian development. Nat Rev Genet 2013;14(3):204–20.

[11] Jones PA, Liang G. Rethinking how DNA methylation patterns are maintained. Nat Rev Genet 2009;10(11):805–11.

[12] Jones PA. Functions of DNA methylation: islands, start sites, gene bodies and beyond. Nat Rev Genet 2012;13(7):484–92.

[13] Cutter A, Hayes JJ. A brief review of nucleosome structure. FEBS Lett 2015;589(20 Pt A):2914–22.

[14] Karlić R, Chung HR, Lasserre J, Vlahovicek K, Vingron M. Histone modification levels are predictive for gene expression. Proc Natl Acad Sci U S A 2010;107(7):2926–31.

[15] Kimura H. Histone modifications for human epigenome analysis. J Hum Genet 2013;58(7):439–45.

[16] Jenuwein T, Allis CD. Translating the histone code. Science 2001;293(5532):1074–80.

[17] Strahl BD, Allis CD. The language of covalent histone modifications. Nature 2000;403(6765):41–5.

[18] Creyghton MP, Cheng AW, Welstead GG, Kooistra T, Carey BW, Steine EJ, et al. Histone H3K27ac separates active from poised enhancers and predicts developmental state. Proc Natl Acad Sci U S A 2010;107(50):21931–6.

[19] Mori A, Deola S, Xumerle L, Mijatovic V, Malerba G, Monsurrò V. Next generation sequencing: new tools in immunology and hematology. Blood Res 2013;48(4):242–9.

[20] Sanger F, Coulson AR. A rapid method for determining sequences in DNA by primed synthesis with DNA polymerase. J Mol Biol 1975;94(3):441–8.

[21] Sanger F, Nicklen S, Coulson AR. DNA sequencing with chain-terminating inhibitors. Proc Natl Acad Sci U S A 1977;74(12):5463–7.

[22] Beck TF, Mullikin JC, NISC Comparative Sequencing Program, Biesecker LG. Systematic evaluation of Sanger validation of next-generation sequencing variants. Clin Chem 2016;62(4):647–54.

[23] Metzker ML. Sequencing technologies—the next generation. Nat Rev Genet 2010;11(1):31–46.

[24] Robinson PN, Piro RM, Jäger M. Computational exome and genome analysis. In: Chapman & Hall/CRC mathematical and computational biology series. Boca Raton, FL: CRC Press (Taylor & Francis Group); 2017.

[25] Lieberman-Aiden E, van Berkum NL, Williams L, Imakaev M, Ragoczy T, Telling A, et al. Comprehensive mapping of long-range interactions reveals folding principles of the human genome. Science 2009;326(5950):289–93.

[26] Wang Z, Gerstein M, Snyder M. RNA-Seq: a revolutionary tool for transcriptomics. Nat Rev Genet 2009;10(1):57–63.

[27] Steijger T, Abril JF, Engström PG, Kokocinski F, RGASP Consortium, Hubbard TJ, et al. Assessment of transcript reconstruction methods for RNA-seq. Nat Methods 2013;10(12):1177–84.

[28] Schadt EE, Turner S, Kasarskis A. A window into third-generation sequencing. Hum Mol Genet 2010;19(R2):R227–40.

[29] Rhoads A, Au KF. PacBio sequencing and its applications. Genomics Proteomics Bioinformatics 2015;13(5):278–89.

[30] Yong WS, Hsu FM, Chen PY. Profiling genome-wide DNA methylation. Epigenetics Chromatin 2016;9:26.

[31] Lu H, Giordano F, Ning Z. Oxford nanopore MinION sequencing and genome assembly. Genomics Proteomics Bioinformatics 2016;14(5):265–79.
[32] Branton D, Deamer DW, Marziali A, Bayley H, Benner SA, Butler T, et al. The potential and challenges of nanopore sequencing. Nat Biotechnol 2008;26(10):1146–53.
[33] Kasianowicz JJ, Brandin E, Branton D, Deamer DW. Characterization of individual polynucleotide molecules using a membrane channel. Proc Natl Acad Sci U S A 1996;93(24):13770–3.
[34] Laszlo AH, Derrington IM, Brinkerhoff H, Langford KW, Nova IC, Samson JM, et al. Detection and mapping of 5-methylcytosine and 5-hydroxymethylcytosine with nanopore MspA. Proc Natl Acad Sci U S A 2013;110(47):18904–9.
[35] Simpson JT, Workman RE, Zuzarte PC, David M, Dursi LJ, Timp W. Detecting DNA cytosine methylation using nanopore sequencing. Nat Methods 2017;14(4):407–10.
[36] Roberts RJ, Carneiro MO, Schatz MC. The advantages of SMRT sequencing. Genome Biol 2013; 14(7):405.
[37] Barros-Silva D, Marques CJ, Henrique R, Jerónimo C. Profiling DNA methylation based on next-generation sequencing approaches: new insights and clinical applications. Genes (Basel) 2018; 9(9):429.
[38] Lister R, O'Malley RC, Tonti-Filippini J, Gregory BD, Berry CC, Millar AH, et al. Highly integrated single-base resolution maps of the epigenome in Arabidopsis. Cell 2008;133(3):523–36.
[39] Frommer M, McDonald LE, Millar DS, Collis CM, Watt F, Grigg GW, et al. A genomic sequencing protocol that yields a positive display of 5-methylcytosine residues in individual DNA strands. Proc Natl Acad Sci U S A 1992;89(5):1827–31.
[40] Meissner A, Gnirke A, Bell GW, Ramsahoye B, Lander ES, Jaenisch R. Reduced representation bisulfite sequencing for comparative high-resolution DNA methylation analysis. Nucleic Acids Res 2005;33(18):5868–77.
[41] Olova N, Krueger F, Andrews S, Oxley D, Berrens RV, Branco MR, et al. Comparison of whole-genome bisulfite sequencing library preparation strategies identifies sources of biases affecting DNA methylation data. Genome Biol 2018;19(1):33.
[42] Wreczycka K, Gosdschan A, Yusuf D, Grüning B, Assenov Y, Akalin A. Strategies for analyzing bisulfite sequencing data. J Biotechnol 2017;261:105–15.
[43] Fraga MF, Esteller M. DNA methylation: a profile of methods and applications. BioTechniques 2002;33(3):632–49.
[44] Huang Y, Pastor WA, Shen Y, Tahiliani M, Liu DR, Rao A. The behaviour of 5-hydroxymethylcytosine in bisulfite sequencing. PLoS One 2010;5(1):e8888.
[45] Booth MJ, Branco MR, Ficz G, Oxley D, Krueger F, Reik W, et al. Quantitative sequencing of 5-methylcytosine and 5-hydroxymethylcytosine at single-base resolution. Science 2012; 336(6083):934–7.
[46] Yu M, Hon GC, Szulwach KE, Song CX, Zhang L, Kim A, et al. Base-resolution analysis of 5-hydroxymethylcytosine in the mammalian genome. Cell 2012;149(6):1368–80.
[47] Yu M, Hon GC, Szulwach KE, Song CX, Jin P, Ren B, et al. Tet-assisted bisulfite sequencing of 5-hydroxymethylcytosine. Nat Protoc 2012;7(12):2159–70.
[48] O'Geen H, Echipare L, Farnham PJ. Using ChIP-seq technology to generate high-resolution profiles of histone modifications. Methods Mol Biol 2011;791:265–86.
[49] Jothi R, Cuddapah S, Barski A, Cui K, Zhao K. Genome-wide identification of in vivo protein-DNA binding sites from ChIP-Seq data. Nucleic Acids Res 2008;36(16):5221–31.
[50] Rhee HS, Pugh BF. Comprehensive genome-wide protein-DNA interactions detected at single-nucleotide resolution. Cell 2011;147(6):1408–19.
[51] Fullwood MJ, Liu MH, Pan YF, Liu J, Xu H, Mohamed YB, et al. An oestrogen-receptor-alpha-bound human chromatin interactome. Nature 2009;462(7269):58–64.
[52] Pajoro A, Muiño JM, Angenent GC, Kaufmann K. Profiling nucleosome occupancy by MNase-seq: experimental protocol and computational analysis. Methods Mol Biol 1675;2018:167–81.
[53] Crawford GE, Holt IE, Whittle J, Webb BD, Tai D, Davis S, et al. Genome-wide mapping of DNase hypersensitive sites using massively parallel signature sequencing (MPSS). Genome Res 2006; 16(1):123–31.

[54] Boyle AP, Davis S, Shulha HP, Meltzer P, Margulies EH, Weng Z, et al. High-resolution mapping and characterization of open chromatin across the genome. Cell 2008;132(2):311–22.
[55] Benichou J, Ben-Hamo R, Louzoun Y, Efroni S. Rep-Seq: uncovering the immunological repertoire through next-generation sequencing. Immunology 2012;135(3):183–91.
[56] Rouet R, Jackson KJL, Langley DB, Christ D. Next-generation sequencing of antibody display repertoires. Front Immunol 2018;9:118.
[57] Workman RE, Tang AD, Tang PS, Jain M, Tyson JR, Zuzarte PC, et al. Nanopore native RNA sequencing of a human poly(A) transcriptome. bioRxiv 2018. bioRxiv 459529.
[58] Adusumalli S, Mohd Omar MF, Soong R, Benoukraf T. Methodological aspects of whole-genome bisulfite sequencing analysis. Brief Bioinform 2015;16(3):369–79.
[59] Chatterjee A, Stockwell PA, Rodger EJ, Morison IM. Comparison of alignment software for genome-wide bisulphite sequence data. Nucleic Acids Res 2012;40(10):e79.
[60] Guo Y, Ye F, Sheng Q, Clark T, Samuels DC. Three-stage quality control strategies for DNA re-sequencing data. Brief Bioinform 2014;15(6):879–89.
[61] Cock PJ, Fields CJ, Goto N, Heuer ML, Rice PM. The Sanger FASTQ file format for sequences with quality scores, and the Solexa/Illumina FASTQ variants. Nucleic Acids Res 2010;38(6):1767–71.
[62] Andrews S. FastQC: a quality control tool for high throughput sequence data, http://www.bioinformatics.babraham.ac.uk/projects/fastqc/; 2010. [Accessed 06 June 2019].
[63] Zhang T, Luo Y, Liu K, Pan L, Zhang B, Yu J, et al. BIGpre: a quality assessment package for next-generation sequencing data. Genomics Proteomics Bioinformatics 2011;9(6):238–44.
[64] Martínez-Alcántara A, Ballesteros E, Feng C, Rojas M, Koshinsky H, Fofanov VY, et al. PIQA: pipeline for Illumina G1 genome analyzer data quality assessment. Bioinformatics 2009; 25(18):2438–9.
[65] Sun S, Noviski A, Yu X. MethyQA: a pipeline for bisulfite-treated methylation sequencing quality assessment. BMC Bioinformatics 2013;14:259.
[66] Lin X, Sun D, Rodriguez B, Zhao Q, Sun H, Zhang Y, et al. BSeQC: quality control of bisulfite sequencing experiments. Bioinformatics 2013;29(24):3227–9.
[67] De Coster W, D'Hert S, Schultz DT, Cruts M, Van Broeckhoven C. NanoPack: visualizing and processing long-read sequencing data. Bioinformatics 2018;34(15):2666–9.
[68] Lanfear R, Schalamun M, Kainer D, Wang W, Schwessinger B. MinIONQC: fast and simple quality control for MinION sequencing data. Bioinformatics 2019;35(3):523–5.
[69] Li H, Durbin R. Fast and accurate short read alignment with Burrows-Wheeler transform. Bioinformatics 2009;25(14):1754–60.
[70] Li H, Durbin R. Fast and accurate long-read alignment with Burrows-Wheeler transform. Bioinformatics 2010;26(5):589–95.
[71] Li H. Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM. arXiv 2013. arXiv:303.3997 [q–bio.GN].
[72] Langmead B, Salzberg SL. Fast gapped-read alignment with Bowtie 2. Nat Methods 2012;9(4):357–9.
[73] Li R, Yu C, Li Y, Lam TW, Yiu SM, Kristiansen K, et al. SOAP2: an improved ultrafast tool for short read alignment. Bioinformatics 2009;25(15):1966–7.
[74] Conesa A, Madrigal P, Tarazona S, Gomez-Cabrero D, Cervera A, McPherson A, et al. A survey of best practices for RNA-seq data analysis. Genome Biol 2016;17(1):13.
[75] Sun X, Han Y, Zhou L, Chen E, Lu B, Liu Y, et al. A comprehensive evaluation of alignment software for reduced representation bisulfite sequencing data. Bioinformatics 2018;34(16):2715–23.
[76] Dobin A, Davis CA, Schlesinger F, Drenkow J, Zaleski C, Jha S, et al. STAR: ultrafast universal RNA-seq aligner. Bioinformatics 2013;29(1):15–21.
[77] Smith AD, Chung WY, Hodges E, Kendall J, Hannon G, Hicks J, et al. Updates to the RMAP short-read mapping software. Bioinformatics 2009;25(21):2841–2.
[78] Xi Y, Li W. BSMAP: whole genome bisulfite sequence MAPping program. BMC Bioinformatics 2009;10:232.
[79] Hoffmann S, Otto C, Kurtz S, Sharma CM, Khaitovich P, Vogel J, et al. Fast mapping of short sequences with mismatches, insertions and deletions using index structures. PLoS Comput Biol 2009;5(9):e1000502.

[80] Pedersen B, Hsieh TF, Ibarra C, Fischer RL. MethylCoder: software pipeline for bisulfite-treated sequences. Bioinformatics 2011;27(17):2435–6.

[81] Chen PY, Cokus SJ, Pellegrini M. BS Seeker: precise mapping for bisulfite sequencing. BMC Bioinformatics 2010;11:203.

[82] Guo W, Fiziev P, Yan W, Cokus S, Sun X, Zhang MQ, et al. BS-Seeker2: a versatile aligning pipeline for bisulfite sequencing data. BMC Genomics 2013;14:774.

[83] Harris EY, Ponts N, Levchuk A, Roch KL, Lonardi S. BRAT: bisulfite-treated reads analysis tool. Bioinformatics 2010;26(4):572–3.

[84] Harris EY, Ponts N, Le Roch KG, Lonardi S. BRAT-BW: efficient and accurate mapping of bisulfite-treated reads. Bioinformatics 2012;28(13):1795–6.

[85] Pedersen BS, Eyring K, De S, Yang IV, Schwartz DA. Fast and accurate alignment of long bisulfite-seq reads. arXiv 2014. arXiv:1401.1129v2 [q-bio.GN].

[86] Krueger F, Andrews SR. Bismark: a flexible aligner and methylation caller for Bisulfite-Seq applications. Bioinformatics 2011;27(11):1571–2.

[87] Li H. Minimap2: pairwise alignment for nucleotide sequences. Bioinformatics 2018;34(18):3094–100.

[88] Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, et al. The sequence alignment/map format and SAMtools. Bioinformatics 2009;25(16):2078–9.

[89] The SAM/BAM Format Specification Working Group. Sequence Alignment/Map Format Specification, version 1.6, https://samtools.github.io/hts-specs/SAMv1.pdf; 2019. [Accessed 06 June 2019].

[90] Van der Auwera GA, Carneiro MO, Hartl C, Poplin R, Del Angel G, Levy-Moonshine A, et al. From FastQ data to high confidence variant calls: the genome analysis toolkit best practices pipeline. Curr Protoc Bioinformatics 2013;43:11.10.1–33.

[91] Laird PW. Principles and challenges of genomewide DNA methylation analysis. Nat Rev Genet 2010;11(3):191–203.

[92] Akalin A, Kormaksson M, Li S, Garrett-Bakelman FE, Figueroa ME, Melnick A, Mason CE. methylKit: a comprehensive R package for the analysis of genome-wide DNA methylation profiles. Genome Biol 2012;13(10):R87.

[93] Benoukraf T, Wongphayak S, Hadi LH, Wu M, Soong R. GBSA: a comprehensive software for analysing whole genome bisulfite sequencing data. Nucleic Acids Res 2013;41(4):e55.

[94] Liang F, Tang B, Wang Y, Wang J, Yu C, Chen X, et al. WBSA: web service for bisulfite sequencing data analysis. PLoS One 2014;9(1):e86707.

[95] Suzuki Y, Korlach J, Turner SW, Tsukahara T, Taniguchi J, Qu W, et al. AgIn: measuring the landscape of CpG methylation of individual repetitive elements. Bioinformatics 2016;32(19):2911–9.

[96] Morales-Nebreda L, McLafferty FS, Singer BD. DNA methylation as a transcriptional regulator of the immune system. Transl Res 2019;204:1–18.

[97] Smith AK, Conneely KN, Kilaru V, Mercer KB, Weiss TE, Bradley B, et al. Differential immune system DNA methylation and cytokine regulation in post-traumatic stress disorder. Am J Med Genet B Neuropsychiatr Genet 2011;156B(6):700–8.

[98] Mack SC, Witt H, Piro RM, Gu L, Zuyderduyn S, Stütz AM, et al. Epigenomic alterations define lethal CIMP-positive ependymomas of infancy. Nature 2014;506(7489):445–50.

[99] Jeschke J, Bizet M, Desmedt C, Calonne E, Dedeurwaerder S, Garaud S, et al. DNA methylation–based immune response signature improves patient diagnosis in multiple cancers. J Clin Invest 2017;127(8):3090–102.

[100] Song Q, Decato B, Hong EE, Zhou M, Fang F, Qu J, et al. A reference methylome database and analysis pipeline to facilitate integrative and comparative epigenomics. PLoS One 2013;8(12):e81148.

[101] Burger L, Gaidatzis D, Schübeler D, Stadler MB. Identification of active regulatory regions from DNA methylation data. Nucleic Acids Res 2013;41(16):e155.

[102] Stadler MB, Murr R, Burger L, Ivanek R, Lienert F, Schöler A, et al. DNA-binding factors shape the mouse methylome at distal regulatory regions. Nature 2011;480(7378):490–5.

[103] Bailey T, Krajewski P, Ladunga I, Lefebvre C, Li Q, Liu T, et al. Practical guidelines for the comprehensive analysis of ChIP-seq data. PLoS Comput Biol 2013;9(11):e1003326.

[104] Thomas R, Thomas S, Holloway AK, Pollard KS. Features that define the best ChIP-seq peak calling algorithms. Brief Bioinform 2017;18(3):441–50.

[105] Zhang Y, Liu T, Meyer CA, Eeckhoute J, Johnson DS, Bernstein BE, et al. Model-based analysis of ChIP-Seq (MACS). Genome Biol 2008;9(9):R137.

[106] Feng J, Liu T, Qin B, Zhang Y, Liu XS. Identifying ChIP-seq enrichment using MACS. Nat Protoc 2012;7(9):1728–40.

[107] Kumar V, Muratani M, Rayan NA, Kraus P, Lufkin T, Ng HH, et al. Uniform, optimal signal processing of mapped deep-sequencing data. Nat Biotechnol 2013;31(7):615–22.

[108] Xing H, Mo Y, Liao W, Zhang MQ. Genome-wide localization of protein-DNA binding and histone modification by a Bayesian Change-Point method with ChIP-seq data. PLoS Comput Biol 2012;8(7): e1002613.

[109] Harmanci A, Rozowsky J, Gerstein M. MUSIC: identification of enriched regions in ChIP-Seq experiments using a mappability-corrected multiscale signal processing framework. Genome Biol 2014;15(10):474.

[110] Steinhauser S, Kurzawa N, Eils R, Herrmann C. A comprehensive comparison of tools for differential ChIP-seq analysis. Brief Bioinform 2016;17(6):953–66.

[111] Landt SG, Marinov GK, Kundaje A, Kheradpour P, Pauli F, Batzoglou S, et al. ChIP-seq guidelines and practices of the ENCODE and modENCODE consortia. Genome Res 2012; 22(9):1813–31.

[112] Bao Y, Vinciotti V, Wit E, PAC 'tH. Accounting for immunoprecipitation efficiencies in the statistical analysis of ChIP-seq data. BMC Bioinformatics 2013;14:169.

[113] Knight JC. Genomic modulators of the immune response. Trends Genet 2013;29(2):74–83.

[114] Tappeiner E, Finotello F, Charoentong P, Mayer C, Rieder D, Trajanoski Z. TIminer: NGS data mining pipeline for cancer immunology and immunotherapy. Bioinformatics 2017;33(19):3140–1.

[115] Love MI, Huber W, Anders S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. Genome Biol 2014;15(12):550.

[116] Kakaradov B, Xiong HY, Lee LJ, Jojic N, Frey BJ. Challenges in estimating percent inclusion of alternatively spliced junctions from RNA-seq data. BMC Bioinformatics 2012;13(Suppl 6):S11.

[117] Ding L, Rath E, Bai Y. Comparison of alternative splicing junction detection tools using RNA-seq data. Curr Genomics 2017;18(3):268–77.

[118] Braunschweig U, Barbosa-Morais NL, Pan Q, Nachman EN, Alipanahi B, Gonatopoulos-Pournatzis T, et al. Widespread intron retention in mammals functionally tunes transcriptomes. Genome Res 2014;24(11):1774–86.

[119] Castel SE, Levy-Moonshine A, Mohammadi P, Banks E, Lappalainen T. Tools and best practices for data processing in allelic expression analysis. Genome Biol 2015;16(1):195.

[120] Rhee JK, Lee S, Park WY, Kim YH, Kim TM. Allelic imbalance of somatic mutations in cancer genomes and transcriptomes. Sci Rep 2017;7(1):1653.

[121] Kim YJ, Kim TH. Chromosome conformation capture for research on innate antiviral immunity. Methods Mol Biol 1656;2017:195–208.

[122] Burren OS, Rubio García A, Javierre B-M, Rainbow DB, Cairns J, Cooper NJ, et al. Chromosome contacts in activated T cells identify autoimmune disease candidate genes. Genome Biol 2017; 18(1):165.

[123] Meagher RB, Müssar KJ. The influence of DNA sequence on epigenome-induced pathologies. Epigenetics Chromatin 2012;5(1):11.

[124] Lo P-K, Zhou Q. Emerging techniques in single-cell epigenomics and their applications to cancer research. J Clin Genom 2018;1:1.

[125] Vegh P, Haniffa M. The impact of single-cell RNA sequencing on understanding the functional organization of the immune system. Brief Funct Genomics 2018;17(4):265–72.

[126] Kunz DJ, Gomes T, James KR. Immune cell dynamics unfolded by single-cell technologies. Front Immunol 2018;9:1435.

[127] Macaulay IC, Ponting CP, Voet T. Single-cell multiomics: multiple measurements from single cells. Trends Genet 2017;33(2):155–68.

This page intentionally left blank

CHAPTER 8

Advances in single-cell epigenomics of the immune system

Jonas Schulte-Schrepping[a], **Humberto J. Ferreira**[b], **Adem Saglam**[b], **Emily Hinkley**[b], **Joachim L. Schultze**[a,b]
[a]Genomics and Immunoregulation, Life & Medical Sciences (LIMES) Institute, University of Bonn, Bonn, Germany
[b]Platform for Single Cell Genomics and Epigenomics at the German Center for Neurodegenerative Diseases and the University of Bonn, Bonn, Germany

Contents

Introduction 185
Single-cell epigenomics technologies 188
  DNA modifications 192
  Protein-DNA interaction 193
Chromatin structure 193
  Chromosome conformation 195
  Single-cell multiomics 195
Computational challenges and solutions 196
  Quality control and preprocessing 198
  Downstream analysis 200
Studying the immune system using single-cell epigenomics 202
  Hematopoiesis 203
  Leukemia 205
  Aging 206
Conclusions and future perspectives 208
  Single-cell epigenomics with a spatial resolution 208
  Other future applications 208
References 209

Introduction

Epigenetics comprises the dynamic molecular layers that regulate the functional accessibility to the genetic information within a cell in processes such as transcription, replication, recombination, and DNA repair [1]. Consequently, it defines lineage- and cell-type specific patterns regulating the transcriptional output during differentiation and development [2, 3]. Over the last few decades, the development of diverse technologies detecting enzymatic genome-wide modifications of DNA nucleotides and histones as well as chromatin structures [4–6] has enabled numerous studies to elucidate heritable epigenetic patterns in both physiological and pathological conditions [7, 8]. For instance, epigenetic fingerprints derived from defined tumor types were shown to robustly classify tumors of

Epigenetics of the Immune System
https://doi.org/10.1016/B978-0-12-817964-2.00008-3

© 2020 Elsevier Inc.
All rights reserved.

unknown origin and could thereby guide treatment selection and improve cancer therapies highlighting their potential clinical relevance [9]. Moreover, epigenetic profiles were shown to improve predictions of a cell's identity compared to corresponding transcriptome profiles [10, 11].

Determining cellular information from bulk samples involves homogenization of signals across potentially heterogeneous populations; thus, molecular patterns of less represented cell types or states are averaged out and eventually lost. Isolation of known cell types or otherwise defined subsets of cells using flow cytometry based on the expression of defined marker genes or microscopic selection of cells based on histological features partially addresses this issue but relies on a priori cell-type definitions. Therefore, a comprehensive investigation of epigenetic heterogeneity among cell populations and determination of cell-type-specific epigenetic regulation across conditions requires methods with single-cell resolution. In recent years, numerous such technologies have been developed to explore chromatin accessibility, DNA methylation, histone modifications, or transcription factor DNA interactions on a single-cell level (Table 1) [12, 13, 15–17].

Table 1 List of publications reporting on single-cell epigenomic technologies.

Date of Publication	Technology	Molecular targets	Reported cells	Reference
2013-09-06	Single-cell DMR-methylation assay	Methylation	80	[12]
2013-09-25	scHi-C	Chromosome conformation	10	[13]
2013-10-31	scRRBS	Methylation	8	[14]
2014-07-20	scBS-seq	Methylation	44	[15]
2015-05-22	sciATAC-seq	Open chromatin	~15,000	[16]
2015-10-12	Drop-ChIP	Histone modifications	5405	[17]
2015-11-05	scHi-C	Chromosome conformation	10	[18]
2016-01-11	scM&T-seq	Methylation and transcriptome	77	[19]
2016-02-23	scTrio-seq	Methylation and transcriptome and copy number variations	37	[20]
2016-05-05	scMT-seq	Methylation and transcriptome	15	[21]
2016-06-23	C1-scATAC	Open chromatin	254	[22]
2016-06-27	scABa-seq	Methylation	480	[23]
2016-08-15	scATAC-seq	Open chromatin	350	[10]
2017-01-30	sciHi-C	Chromosome conformation	10,696	[24]
2017-03-29	snHi-C	Chromosome conformation	120	[25]
2017-06-27	scNOME-seq	Open chromatin and methylation	30	[26]
2017-08-11	snmC-seq	Methylation	>6000	[27]
2017-12-07	MscRRBS	Methylation and transcriptome	706	[28]
2017-12-11	scTHS-seq	Open chromatin and transcriptome	~60,000	[29]

Table 1 List of publications reporting on single-cell epigenomic technologies—cont'd

Date of Publication	Technology	Molecular targets	Reported cells	Reference
2018-02-22	scNMT-seq	Open chromatin and methylation and transcriptome	61	[30]
2018-03-31	scATAC-seq	Open chromatin and transcriptome	2034	[31]
2018-04-09	sci-MET	Methylation	3282	[32]
2018-08-23	sciATAC-seq	Open chromatin	>100,000	[33]
2018-09-07	scATAC-seq	Open chromatin	3949	[34]
2018-09-20	snmC-seq2	Methylation	3072	[35]
2018-09-28	sci-CAR	Open chromatin and transcriptome	16,121	[36]
2018-11-02	Pi-ATAC	Open chromatin	1559	[37]
2018-12-17	plate-based scATAC-seq	Open chromatin	3688	[38]
2018-12-26	sn-m3C-seq	Chromosome conformation	248	[39]
2018-12-28	iscCOOL-seq	Open chromatin and methylation	667	[40]
2019-01-28	scCAT-seq	Open chromatin and transcriptome	74	[41]
2019-02-11	scM&T-seq	Methylation and transcriptome	186	[42]
2019-03-28	scChIC-seq	Histone modifications	242	[43]
2019-04-01	sci-ATAC-seq	Open chromatin	2346	[44]
2019-04-18	scATAC-seq	Open chromatin	>200,000	[45]
2019-04-29	Cut&Tag	Histone modifications, RNA polymerase II and transcription factor binding	808	[46]
2019-06-24	dsc(i)ATAC-seq	Open chromatin	510,123	[47]
2019-07-07	snATAC-seq	Open chromatin	7846	[48]

Although all currently available NGS-based single-cell epigenomics technologies exhibit fundamental difficulties in signal detection due to the inherently limited material per cell (i.e., two copies of genomic DNA in diploid organisms), they have already been successfully applied to elucidate cellular heterogeneity in diverse biological contexts [10, 17, 31, 49]. In particular, techniques to study chromatin accessibility patterns in individual cells have advanced quickly (Table 1, Fig. 1), arguably due to their relatively simple strategies, and already allow high-throughput analysis of hundreds of thousands of cells in a single experiment [47]. Simultaneously, computational methods have been and are actively being developed to address the challenging demands of this new data type and allow for efficient processing and statistically sound interpretations [50–52].

Constant epigenetic programming during hematopoiesis as well as activation-induced proliferation and differentiation makes the immune system a paradigmatic example to study epigenetic regulation on the single-cell level. Moreover, blood is an easily accessible tissue composed of diverse cell populations, thus being a very attractive system to explore epigenetic heterogeneity in humans. Additionally, tissue-resident immune cells, such as macrophages [53] and microglia [29, 33], are known to exhibit complex

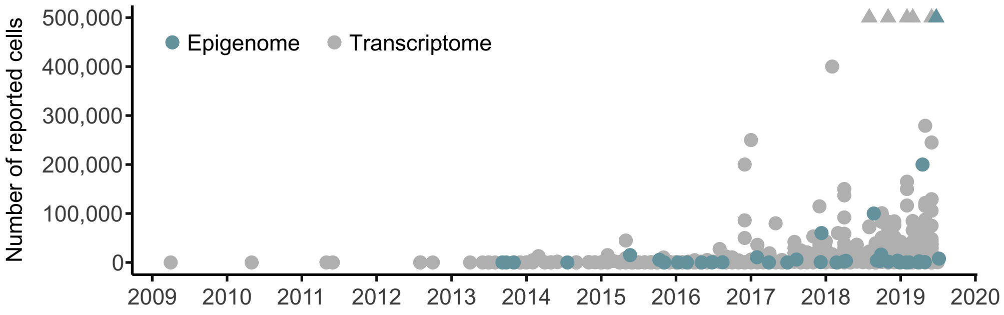


Fig. 1 The evolution of single-cell epigenomics technologies. Depicted is the number of cells reported per publication sorted by date of publication. Following the rapid evolution of scRNA-seq technologies from 9 single cells up to 2,000,000 within 10 years (gray), single-cell epigenomics (turquoise) is evolving similarly, rapidly reaching a current maximum of over 500,000 cells as reported by Lareau et al. [47]. Publications reporting more than 500,000 cells are depicted as triangles on the top edge of the plot.

environment-dependent epigenetic profiles encouraging the use of single-cell epigenomics to elucidate the complete scale of intercellular variability. Here also, spatial resolution becomes increasingly important to assess intra-tissue heterogeneity depending on the cellular (co-)localization and points to potential future developments in the field of single-cell epigenomics.

In this chapter, we describe the most common single-cell epigenomic technologies currently available, including multiomics approaches, as well as discuss challenges of and solutions for computational data analysis and integration. We present selected studies underlining the relevance of such methods to decipher epigenetic heterogeneity with particular emphasis on the immune system.

Single-cell epigenomics technologies

Most next-generation sequencing (NGS)-based single-cell epigenomics methods are adapted from their equivalent bulk protocols (Table 2). Independently of the downstream read out, the main challenges in the conversion of a bulk analysis method to single-cell level is the efficient isolation of individual cells, the limited amount of genomic material per cell, multiplexed processing of cells based on cellular barcoding, as well as the cost per cell.

Technical advances in automated liquid handling and cell sorting devices have had a direct impact on the development and throughput of single-cell sequencing technologies. In any single-cell technique, the initial step consists in the isolation of single cells from cell cultures or tissues [65]. Of course, this can be achieved by manually isolating a single cell with a micropipette under the microscope [66]. However, with this kind of method, the number of cells that can be isolated remains low, which is only useful for rare

Table 2 Overview of bulk and corresponding single-cell epigenomics techniques.

Target	Feature	Bulk method	Single-cell methods	Approach	Reference
DNA modification	5mC/5hmC	BS-seq RRBS-seq	scBS-seq (M) scRRBS-seq	Unmethylated C's are converted into T's, while 5mC's and 5hmC's remain unconverted	[14, 15, 28, 54]
	5mC and 5hmC	EM-seq	–	5mC's and 5hmC's are oxidized and protected from conversion	[55]
	5hmC	Aba-seq	scAba-seq	5hmC's are glucosylated, followed by glucosylation-dependent DNA digestion	[23]
Protein-DNA interaction	Histone modification	ChIP-seq	scChIP-seq	DNA and transcription factors are first cross-linked and then immunoprecipitated	[17]
	Transcription factor binding	DamID	scDamID	Transfection of cells with a fusion of transcription factor gene of interest and Dam gene which produces a protein that methylates adenine residues in proximity of the binding site	[56, 57]
Chromatin structure	Nucleosome positioning	MNase-seq	scMNase-seq	Micrococcal nuclease digests open chromatin regions. Nucleosome regions are isolated and sequenced.	[26, 58]
		NOMe-seq	scNOMe-seq	GpC methyltransferase M.CviPI methylates GpC sites unbound to nucleosomes, followed by bisulfite treatment	[26, 59]
	DNA accessibility	DNase-seq	scDNase-seq	DNaseI enzyme is used to cut accessible double-stranded DNA, which is followed by ligation of known primer sequences	[60]
		ATAC-seq	scATAC-seq sci(p)ATAC-seq dsc(i)	Tn5 transposase is used to cut accessible double-stranded genomic DNA and ligate known primer sequences to the 5′-ends	[16, 31, 33, 47, 61]

Continued

Table 2 Overview of bulk and corresponding single-cell epigenomics techniques—cont'd

Target	Feature	Bulk method	Single-cell methods	Approach	Reference
			ATAC-seq		
		THS-seq	scTHS-seq	Uses a customized Tn5 transposome system that is engineered to attach a T7 promoter to the end of every DNA molecule after fragmentation	[29, 62]
Chromosome conformation	Locus-locus interaction	3C-seq	sc3C-seq	Chromatin is cross-linked with formaldehyde followed by restriction digestion and religation, resulting in chimeric DNA fragments	[63, 64]
		HiC-seq	scHiC-seq sciHiC-seq	Chromatin is cross-linked with formaldehyde followed by restriction digestion, biotin fill-in, ligation of biotinylated ends resulting in chimeric DNA fragments	[13, 24, 63]

cell types or very small pieces of tissues that are difficult to collect from living organisms, such as brain tissues [67].

Therefore, a range of methods have been developed to enable nearly automated, high-throughput cell isolation [68]. Importantly, each of these methods has specific advantages and disadvantages influencing the cells studied. They can thus critically affect data quality and interpretation.

Fluorescent-activated cell sorting (FACS) can be deployed to sort single cells into wells of 96- or 384-well plates. This has the advantage of using the expression of cellular markers to select cell populations and simultaneously record protein expression levels for individual cells, hence called index sorting [69]. Alternatively, microfluidic devices can serve as automated cell dispensers using automated object recognition algorithms [70] to sort single cells into 96- or 384-well plates. In these plate-based methods, library preparation from each cell is performed separately, which improves the sensitivity but limits the throughput and results in high cost per cell. While in bulk techniques thousands to millions of cells are processed together, reducing the de facto reaction volumes per cell to the nanoliter scale, limitations of liquid handling demand certain reaction volumes per cell when processed in isolation, which increases the price per cell dramatically.

In order to increase the throughput of sequencing techniques and reduce the cost per sample or cell, barcodes can be introduced into the sequencing library to allow for early multiplexing for combined processing and sequencing [71]. The design of such barcode sequences is critical for efficient in silico demultiplexing of single-cell libraries at low error rates.

Development of droplet-based microfluidics protocols presented major progress in terms of throughput for single-cell omics applications. Here, single cells are captured together with RNA- or DNA-capture beads in aqueous droplets in a flow of oil through microcapillary channels [17]. Each bead has a unique cell barcode in its coupled primer sequence and thus allows for in silico allocation of reads to individual cells. Unfortunately, all microfluidics-based methods have the disadvantage of putting cells under harsh conditions (high pressure), which can impair the quality of the cells and eventually destroy fragile cells impeding their analysis. The C1 instrument developed by Fluidigm [72], the Chromium Controller by 10X Genomics [73], and the ddSEQ single-cell isolator by Illumina/Bio-Rad are commercially available adaptations of this technique. In contrast, although not yet utilized in single-cell epigenomics protocols, gravity-based capture of single cells and beads in picoliter wells, as for example used in SeqWell [74], has the advantage of reducing shearing forces that might otherwise affect the cells. Importantly, the well size should only allow for accommodation of a single cell and a single barcoded bead. Importantly, in all bead-based technologies, the ratio of beads and cells must be carefully adjusted based on the Poisson distribution and rigorously controlled to prevent doublet formation. Unfortunately, this involves a vast number of "empty" beads aggravating the distinction of true cellular signal from background noise.

Laser capture microdissection (LCM) presents another single-cell isolation method that uses a high-power ultraviolet (UV) laser to cut selected cells from a tissue based on histological features and catapults them into a collector tube or plate. This technique has the additional advantage of tracking the spatial location of each collected cell in the tissue [75, 76], but is limited in throughput.

Although not truly single-cell isolation techniques, combinatorial indexing approaches [16, 77] fulfill the purpose of separating single cells by introducing unique combinations of barcodes into each single-cell library using split and pool strategies. These approaches have the huge advantage that all steps are essentially performed in bulk, thereby drastically reducing the price per cell, but have been shown to be less sensitive.

Common to all single-cell omics technologies independently of the cell isolation approach is the difficulty in detecting reliable signals from the limited source material per cell. While every cell can have several copies of one RNA transcript, genomic DNA is typically present in only two copies in diploid organisms, unless the cell undergoes mitotic or meiotic division or is otherwise altered genetically, for example cancer cells. Therefore, signal detection in individual cells is inherently limited and requires amplification steps to prevent loss of biomolecules during sequencing library preparation and thus improve the chance of detection. However, these amplification steps can cause distorted sequencing libraries due to PCR biases and consequentially inaccurate quantification results [78]. In combination with the limited sensitivity of most single-cell omics techniques, this results in very sparse data with high noise levels making analysis much more challenging than in bulk.

In this section, we present the most commonly applied single-cell epigenomic technologies divided by their biological read-out (Table 1), i.e., DNA modifications, protein-DNA interactions, chromatin structure, and 3D conformation.

DNA modifications

Modified genomic nucleotides, such as methylated (5mC) or hydroxymethylated (5hmC) cytosine residues, are involved in the epigenetic regulation of many important biological processes. The gold-standard method for detection of cytosine methylation at single-base resolution is bisulfite conversion followed by sequencing (BS-seq) [14, 15, 54]. The unmodified method, however, is not able to distinguish between the two types of cytosine methylation. Another limitation of this technique is the rapid degradation of DNA after bisulfite treatment [15], i.e., single-stranded DNA is fragmented at some converted cytosines. Therefore, a big amount of starting material is required for reliable signal detection preventing the application of this technique to genomic DNA from single cells. For the first time, the analysis of cytosine methylation was achieved at the single-cell level using a "reduced-representation bisulfite sequencing" approach (scRRBS-seq), which is based on enrichment of CpG dense regions by using multiple restriction enzymes to produce sequence-specific fragmentation followed by methylated adaptor ligation [14]. However, this technique suffers from poor coverage. Multiplexed scRRBS-seq (MscRRBS-seq) presents an alternative approach, in which single cells are isolated by sorting followed by restriction enzyme digestion and barcoded adaptor ligation [79]. Since every single cell receives a unique barcoded adaptor, all cells can be pooled and treated jointly with bisulfite for efficient CT (cytosine to thymine) conversion. The development of a post-bisulfite adapter-tagging (PBAT) approach, in which the bisulfite conversion step is performed before library preparation so that DNA fragmentation does not destroy adaptor-tagged fragments [80], further improved the detection of genome-wide cytosine methylation yielding a 50% coverage of the CpG methylation in a single cell.

Hydroxymethylated cytosine (5hmC) has been analyzed in bulk samples using modified bisulfite sequencing methods [81, 82], 5hmC-specific restriction enzymes [83], or immunoprecipitation [84]. ScAba-seq [23] can map 5hmC's on a single-cell level by using a glucosylation-dependent DNA digestion. First, 5hmC is glucosylated and the

DNA is further digested using an enzyme that recognizes 5-glucosylhydroxymethylcytosine. The samples are then prepared for sequencing by ligation of adaptor primers. Recently, an alternative method to the bisulfite treatment was developed for the study of 5mC and 5hmC modifications [55]. This method is based on a two-step enzymatic reaction that involves the oxidation of 5mC and 5hmC by TET2, followed by deamination and conversion of non-oxidized cytosine to uracils, while oxidized 5mC and 5hmC bases are protected. This so-called Enzymatic-Methyl-sequencing (EM-seq) technique has several advantages over BS-seq. It results in larger library insert sizes and much less GC bias with improved CpG coverage due to the use of enzymatic processes instead of harsh chemical treatment. However, its single-cell equivalent has not been developed yet.

Protein-DNA interaction

Histone marks—diverse covalent modifications of histones, such as H3K4me3 or H3K27ac, that represent different transcriptional states of a genomic locus—can be mapped by chromatin immunoprecipitation followed by sequencing (ChIP-seq) [17]. High background noise due to nonspecific antibody pull-down initially hampered the application of ChIP-seq at the single-cell level. This was only recently overcome by combining droplet-based microfluidics with genomic barcoding. Although this approach yields very limited numbers of valid sequencing reads per cell due to an inefficient amplification and barcoding strategy, the high number of cells assayed still allows for assessing intercellular variability.

Alternatively, transcription factor-DNA interactions can be assessed using "DNA adenine methyltransferase identification sequencing" (DamID-seq) [57]. The cells are first transfected with a fusion gene encoding for the transcription factor of interest and the DAM gene from *Escherichia coli* (*E. coli*). Upon expression, the Dam protein methylates adenosines in the neighborhood of the binding site of the transcription factor. Methylated sites are then cut by the methylation-sensitive restriction enzyme *Dpn*I followed by ligation of sequencing adaptors. The technique's single-cell adaptation (scDamID-seq) was successfully used to determine the genomic interaction of nuclear lamina protein Lamin B1 [56].

Chromatin structure

The double-stranded genomic DNA located in the nucleus is sustained by histone octamers in a quasi-regular pattern. Each histone octamer is encircled by approximately 146 bp of DNA [85]. This histone-DNA complex is called nucleosome. In regions of the chromatin where nucleosomes are densely packed, i.e., closed chromatin regions, transcription factors are unable to bind and transcription cannot occur as opposed to open chromatin regions where nucleosomes are relaxed. As the conformation of nucleosomes is controlled by a variety of epigenetic modifications, such as DNA methylation and histone modifications, changes in the structural organization of the chromatin indirectly represent epigenetic changes and indicate transcriptionally active or primed genomic regions.

Technically, mapping of open chromatin regions can be carried out in different ways mostly using enzymes that cut accessible regions in the double-stranded genomic DNA. Initially, DNase-seq was developed to study open chromatin regions using DNaseI [60, 86], an enzyme that nonspecifically cleaves DNA free of nucleosomes and transcription factors. The DNase cleavage must be followed by a ligation step to add adaptors to the free ends of the cleaved DNA fragments. Alternatively, FAIRE-seq [87] relies on differences in cross-linking efficiencies between DNA and nucleosomes or sequence-specific DNA-binding proteins. Nucleosome-free regions of DNA are less efficiently cross-linked to proteins and can thus be readily segregated from cross-linked DNA by phenol/chloroform extraction.

Today, the assay for transposase-accessible chromatin (ATAC-seq) [88], in which a Tn5 transposase is used to fragment DNA and simultaneously introduce adapter sequences, is the most widely applied technique. This process of simultaneous fragmentation and tagging is referred to as tagmentation [65, 89]. Open regions of the chromatin can be assessed by introducing the transposase into permeabilized nuclei, where it tagments nucleosome- and transcription factor-free stretches of DNA [88]. Initially, ATAC-seq was adapted to single-cell level using a combinatorial indexing strategy called single combinatorial indexing ATAC followed by sequencing (sciATAC-seq). In this approach, nuclei are tagmented together in multiple pools, thereby introducing a first barcode [16]. These pools are then combined and randomly split again for introduction of a second barcode by polymerase chain reaction (PCR), thereby theoretically producing a unique barcode combination for each cell. In order to keep the doublet rate, i.e., the probability that two single cells get the same barcode combination, sufficiently low, the number of pools and cells per pool must be optimized. Recently, an improved version of this strategy, called scipATAC-seq [61], was developed using the small molecule inhibitor Pitstop 2 [90] to improve penetration of the transposase into the nucleus. In parallel, Buenrostro and colleagues adapted their initial ATAC-seq protocol to single-cell level using microfluidics [22] to capture cells in micro-chambers for lysis and subsequent library preparation. The commercially available microfluidics device initially used for this purpose allowed for parallel tagmentation of 96 single cells. Very recently, droplet-based approaches for scATAC-seq increased the throughput up to hundreds of thousands of single cells processed in parallel [45, 47].

ATAC-seq-based approaches have several inherent drawbacks. First, the Tn5 transposome complex loaded with two different adaptors inserts these in random orientation, resulting in theoretical loss of ca. 50% of molecules during PCR amplification. Second, since approximately only 1% of the genome is accessible in most human cells, regions in which two adjacent tagmentation reactions occur too far apart from each other cannot be amplified by PCR. Finally, short accessible regions experience only a few tagmentation events. These three factors strongly decrease the sensitivity of these ATAC-seq methods. The transposome hypersensitive sites sequencing (THS-seq) assay was developed to overcome these limitations [62]. In THS-seq, a customized Tn5 transposome system is engineered in order to attach a T7 promoter to the end of every DNA molecule after tagmentation. This was later adapted to yield the scTHS-seq method, which is based on a combinatorial indexing approach using barcoded Tn5 adaptors [29].

Investigation of open chromatin regions can also be assayed by identifying nucleosome occupancy. In bulk and single-cell samples, genome-wide nucleosome occupancy profiling has been achieved by sequencing the products resulting from MNase digestion (MNase-seq/scMNase-seq) [26, 58] as well as by nucleosome occupancy and methylome sequencing (NOMe-seq/scNOMe-seq) [26, 59, 91]. In the latter, the native chromatin is treated with the GpC methyltransferase enzyme M.CviPI to methylate exposed GpC dinucleotides while DNA bound by nucleosomes remains unaffected. Mapping of nucleosome positions and methylated sites can then be done by sequencing the bisulfite-converted DNA, which represents a simple but very elegant multiomics approach.

Chromosome conformation

In addition to determining the chromatin landscape, it is also possible to assess the three-dimensional chromosome conformation at the single-cell level using high-throughput chromosome conformation capture followed by sequencing (HiC-seq) [13, 18, 92]. This method is based on Chromosome Conformation Capture (3C), which relies on cross-linking the chromatin with formaldehyde followed by digestion and religation in a way that only covalently linked DNA fragments form ligation products. These fragments are then reverse-cross-linked and detected by qPCR [63, 64]. Among others, the HiC approach extends this strategy to achieve genome-wide detection of chromatin contacts using biotin-dependent enrichment of fragments and high-throughput sequencing. The resulting sequencing library gives information about which genomic regions, although possibly highly separated on the genome, are spatially close to each other [63]. The HiC-seq method was adapted to single-cell level by performing chromatin cross-linking, restriction digestion, biotin fill-in, and ligation within single nuclei. Recently, Ramani and colleagues adapted scHiC-seq by using combinatorial indexing (sciHiC-seq) to obtain libraries of more than 10,000 single cells [77].

Single-cell multiomics

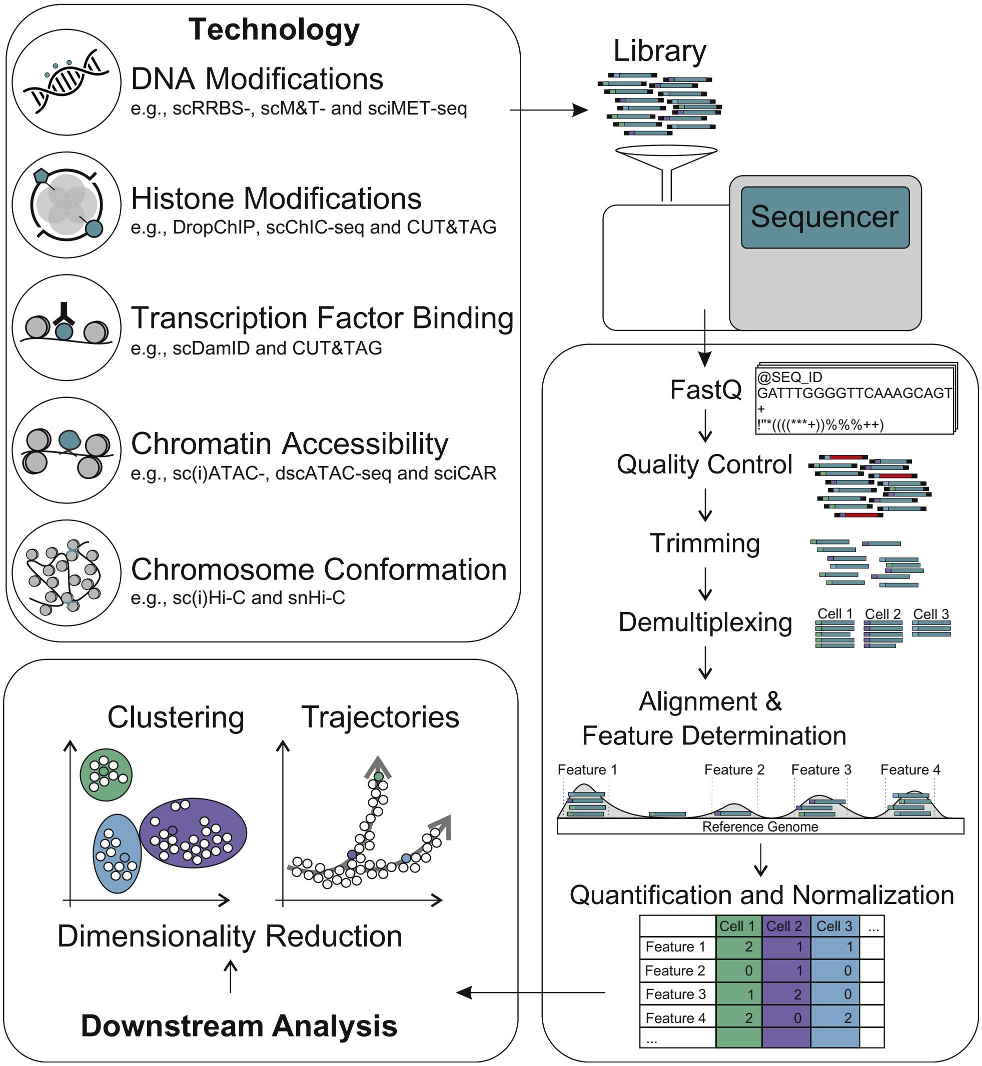
Each single-cell sequencing technique addresses a specific feature of a cell's epigenetic landscape, which gives only partial information on the cellular activity. In order to gain a better understanding of the different mechanisms and pathways that influence, e.g., the activation of immune cells, one needs to gather as much information as possible, which is feasible only by combined application of different single-cell omics technologies.

There has been a lot of effort trying to combine different sequencing methods since each of them requires different conditions. In particular, the combination of transcriptomics and epigenomics was of great interest for the analysis of epigenetic-transcriptional correlations, enabling detailed investigations of how epigenetic patterns are associated with phenotypic changes. Researchers were able to combine BS-seq and RNA-seq to develop the single-cell methylome and transcriptome sequencing technique (scM&T-seq) [19]. This was achieved by physically separating the mRNA from DNA based on genome and transcriptome sequencing or G&T-seq [93], allowing for intricate investigations of links between epigenetic and transcriptional heterogeneity within a particular cell and tissue type. Furthermore, to measure chromatin accessibility together with DNA methylation, NOMe-seq was adapted and incorporated in the scM&T-seq protocol, thereby creating single-cell nucleosome, methylation, and transcription sequencing (scNMT-seq) [30]. Coupling of ATAC-seq or DNase-seq with RNA-seq was also achieved by means of physical separation of RNA and DNA using either combinatorial indexing (sci-CAR) [36] or by parallel amplification [94]. With advances in the different single-cell epigenomic technologies, we expect the range of multiomics approaches to increase dramatically and provide new insights in epigenetic regulation of various biological processes across tissues both in physiological and pathological conditions.

Computational challenges and solutions

The computational analysis of NGS-based omics data can generally be divided into two main parts. The first, referred to as preprocessing, is to transform the vast bulk of sequencing reads into a parsable matrix of biologically classified quantifications, such as chromatin accessibility values, while efficiently detecting and removing obvious error sources. The second is to extract meaningful and statistically sound insights from the resulting high-dimensional count matrix by means of, e.g., dimensionality reduction, sample clustering, and inferential statistics, summarized under the term downstream analysis (Fig. 2).

With the advent of single-cell omics technologies and the concomitant drastic increase in sample size from at most several thousand bulk samples to tens of thousands of single cells (Table 1, Fig. 1), new computational and analytical challenges emerged to fully exploit this data type. As single-cell transcriptomics is leading the advancement of single-cell omics, the majority of analysis software so far has been conceived with the specific objectives and challenges of single-cell gene expression analysis in mind. One major hurdle for applying these algorithms to epigenome data is the divergent nature of the data. Algorithms written for scRNA-seq data expect a data type distinct from epigenome data and thus often utilize inappropriate mathematical models. While single-cell transcriptomics yields continuous gene expression values representing the copy number


Technology
DNA Modifications
e.g., scRRBS-, scM&T- and sciMET-seq
Histone Modifications
e.g., DropChIP, scChIC-seq and CUT&TAG
Transcription Factor Binding
e.g., scDamID and CUT&TAG
Chromatin Accessibility
e.g., sc(i)ATAC-, dscATAC-seq and sciCAR
Chromosome Conformation
e.g., sc(i)Hi-C and snHi-C
Library
Sequencer
FastQ
Quality Control
Trimming
Demultiplexing
Cell 1
Cell 2
Cell 3
Alignment &
Feature Determination
Feature 1
Feature 2
Feature 3
Feature 4
Reference Genome
Quantification and Normalization
Clustering
Trajectories
Dimensionality Reduction
Downstream Analysis

```
@SEQ_ID
GATTTGGGGTTCAAAGCAGT
+
!"*((((***+))%%%++)
```

	Cell 1	Cell 2	Cell 3	...
Feature 1	2	1	1	
Feature 2	0	1	0	
Feature 3	1	2	0	
Feature 4	2	0	2	
...				

Fig. 2 Overview of the epigenomics techniques and a generic analysis pipeline. Depicted is a schematic representation of the most common single-cell epigenomic technologies and their molecular targets as well as a simplified overview of the typical analysis steps covering both the preprocessing of raw sequencing data and the downstream analysis.

of mRNA molecules present in a cell at a certain time, epigenomics produces nearly binary data in the sense that a specific region of each copy of genomic DNA in a single cell of a diploid organism is, e.g., either accessible or not, resulting in values of 0, 1, or 2. However, dedicated software for handling single-cell epigenomics data is actively being developed.

Data produced by currently available single-cell epigenomic protocols suffer from high sparsity and noise. Deriving biologically meaningful interpretations from such data requires elaborated analysis strategies. In this section, we will outline the general steps for analyzing data produced by the most widely applied NGS-based single-cell epigenomics protocols (Fig. 2), discuss potential pitfalls and difficulties, and highlight selected software packages. So far, most studies applying single-cell epigenomics technologies rely on custom scripts and experiment-specific analysis solutions. Nevertheless, the first software suites to facilitate and standardize the preprocessing of single-cell epigenomics data have been developed and provided to the community, such as *proatac* (http://buenrostrolab.com/proatac/). Additionally, several user-friendly software packages, such as *SCRAT* [95], *ScAsAt* [96], or epiScanpy [97], combine multiple functionalities for either or both preprocessing and comprehensive downstream analysis of single-cell ATAC-seq data in one toolkit. As single-cell ATAC-seq is currently most advanced among single-cell epigenome technologies, the following description is biased toward this technology but intends to touch upon all techniques.

Quality control and preprocessing

Single-cell epigenomics is still in its infancy. Multiple steps in the complex protocols, such as sample procurement, tissue preservation, and processing or single-cell isolation, can critically affect the data quality and have not yet been systematically evaluated and standardized. Thus, the first step of analysis should always be quality control including the base-calling quality, the sequence duplication rate, or the presence of common sequencing adapters. In case low-quality bases or sequencing adapter contaminations are observed, trimming is necessary to enable best possible alignment against a reference genome, which can be performed using, e.g., *Trimmomatic* [98].

Depending on the experimental protocol for library production, demultiplexing—the process of sorting and allocating sequencing reads to their cell of origin—has to be achieved on different levels. While in early plate-based techniques single cells are processed separately and indexed only for joint sequencing, recently developed massive throughput techniques [47] deploy early introduction of cell-specific barcodes or combinations thereof [16] for immediate multiplexing and joint processing.

Next, the genomic sequences of the reads are aligned against a reference genome to determine their genomic origin and allow for quantification with genomic resolution using, e.g., *Bowtie 2* [99]. Subsequently, the aligned reads are cleaned, e.g., from PCR duplicates or those mapping to genomic regions known to produce high amounts of unspecific signal, such as the mitochondrial genome [100]. Additional data processing steps might be necessary depending on the specific assay, such as shifting the read position by 9 bp due to the Tn5 adapter insertion bias in ATAC-seq [88].

In contrast to transcriptome analyses, in which quantification of gene expression largely relies on predefined gene models, there is no such reference annotation for epigenetic features across tissues, cell types, or states. A simple solution for feature definition in epigenome analyses is to partition the genome in windows of arbitrary width or extended gene regions for quantification of the read coverage. Alternatively, the exact regions of open chromatin (ATAC-seq, DNase-seq, NOME-seq), histone methylation (BS-seq), or other histone modifications and transcription factor DNA interactions (CHIP-seq) must be derived directly from the sequencing reads by distinction of signal accumulation from background noise, referred to as peak calling—a nontrivial process in light of the low signal-to-noise ratio of single-cell epigenome profiles. For this purpose, reads from all single-cell libraries are typically combined and used as a pseudo-bulk sample to efficiently define the feature space for subsequent per-cell signal quantification. *MACS2* is a widely used peak caller [101]. In order to put the identified features into a genetic context, peaks are usually annotated based on their localization to nearby genes. Although inherently less informative than sample-specific peaks, genomic windows have the substantial advantage to enable direct comparison and integration of signal from different samples across experiments without reprocessing.

Once the set of genomic features has been determined, the signal from each cell is quantified and combined in a matrix of counts per feature (rows) per cell (columns). Depending on the respective genomic feature, appropriate normalization presents the next critical step to prevent technical differences from affecting downstream analyses. In case of genomic windows or extended gene body regions, counts can be normalized for sequencing depth, ergo the total number of reads detected in all features per sample, scaled and log-transformed. This normalization approach, however, violates the assumption of binary signals quantified from defined peak regions, which requires adapted normalization strategies depending on the performed downstream analysis [51, 96, 102].

The high sparsity in combination with the large sample size suggests accounting for missing information based on technical dropouts by means of pooling or imputation. Pooling information across cells with similar profiles is one strategy to reduce zeros and enrich the data at the cost of averaging signal and obscuring heterogeneity, referred to as pseudo-bulk analysis [27]. On the other hand, imputation of missing information using predictive models based on machine learning has been introduced recently for single-cell DNA methylation data [103].

Especially in high-throughput combinatorial-indexing or bead-based microfluidics protocols [16, 47], in which cell barcodes present an arbitrary selection of random sequences and the number of detected cells is unknown, a major challenge lies in the discrimination of true single-cell signals from doublets or background noise. A typical metric guiding this decision is the number of fragments overlapping previously defined peak regions per cell.

Single-cell hi-C protocols require a dedicated analysis strategy as both ends of the sequencing reads represent distant genomic locations and the data suffer from very high noise levels. Shortly after demultiplexing and trimming, reads from each pair are aligned separately and strictly filtered based on various quality metrics, including their frequency in the sequencing library and their proximity to corresponding restriction sites. Subsequently, distinct read pairs are identified and the coverage of the putative contacts is quantified and normalized to reconstruct the 3D genome organization for each cell and compare respective changes across conditions [13, 77, 104, 105].

Downstream analysis

Once the preprocessing and data transformation of raw sequencing reads into a biologically interpretable matrix has been completed, diverse strategies can be followed to extract meaningful insights. Here, the epigenetic modification measured as well as the biological question at hand directs the downstream analysis. For example, investigating epigenetic changes in histone modifications throughout developmental processes or cellular differentiation demands a different strategy than cell-type classification or subtype discovery based on chromatin accessibility profiles. Common to all analyses of high-dimensional single-cell omics data whatsoever is the aim to capture the overall data structure by condensing the high-dimensional feature space down to its most relevant components. The resulting lower dimensional representation exhibits significantly less noise and allows for comprehensible visualization as well as clustering of similar cells. Typically, principal component analysis (PCA) can be applied in order to reduce the number of variables while preserving as much information as possible. However, the high sparsity and noise level of the counts in single-cell epigenomic data drastically affect the performance of this and other conventional methods. Therefore, different strategies have been described to improve dimensionality reduction methods for single-cell epigenomic data. Topic modeling approaches developed for natural language processing [106] have been successfully applied for data transformation before PCA or to directly identify sets of features that share similar accessibility patterns, thereby reducing the feature space [38, 51, 107]. Dimensionality reduction based on cell-by-cell Jaccard distances presents another approach as used for example in ScAsAT [96] or SnapATAC [102]. Moreover, Destin [108] improves PCA by using peak-specific weights calculated based on the distances to transcription start sites (TSSs) as well as the relative frequency of chromatin accessibility peaks in ENCODE reference data.

Alternatively, a group of novel methods do not work on signals derived directly from genomic features, such as genomic windows or predefined peaks, but use biological metafeatures instead to produce a reduced dimensionality representation. SCRAT [95] and chromVAR [52], for example, aggregate read counts across biological features, such as transcription factor binding motifs, reference peaks, or gene sets of interest.

Cicero [51] infers gene activity scores from single-cell open chromatin profiles by identifying correlated regulatory elements and linking these to target genes using unsupervised machine learning.

Independent of the strategy used for initial dimensionality reduction, recently developed machine learning methods, e.g., t-distributed stochastic neighbor embedding (t-SNE) [109] or Uniform Manifold Approximation and Projection (UMAP) [110] can be used to project the resulting data into a 2D space for visualization.

As single-cell epigenomics protocols are complex and far from standardized, technical artifacts, typically referred to as batch effects, can dominate the data structure and hamper biological interpretation in case experiments were performed separately in time and space. For example, Baker et al. reported obvious batch effects between open-chromatin profiles of the same cells obtained from separate runs and solved the issue by removing features detected only in single batches [96]. Although not reported so far, longitudinal or multisite studies applying single-cell epigenomics protocols in more complex settings will demand the development of more sophisticated solutions following examples from scRNA-seq [111], as removing sample-specific signals might involve loss of biologically relevant information.

Depending on the strategy used for dimensionality reduction, appropriate clustering methods are applied to infer similarities between cells and determine groups of cells potentially representing distinct cell types or specific states [17, 51, 52]. Consequently, cluster-specific signals can be determined to characterize the cluster identities and define biologically distinct populations [33, 47]. Additionally, annotation of single-cell profiles according to known cell types helps to validate clusters and locate the data at hand in the universe of diverse cellular identities. For example, SCRAT [95] enables the comparison of each cell's chromatin profile to a database of ENCODE DNase-seq profiles from a wide variety of cell types. Obviously, the performance of such reference-based strategies strictly depends on the quality of the reference data and does not allow for de novo classification of novel subtypes.

As the epigenome regulates a cell's potential to differentiate, single-cell profiles spanning multiple stages enable the reconstruction of continuous differentiation trajectories, which are not always well described by discrete clusters. STREAM [112] presents a comprehensive software pipeline to disentangle and visualize branching trajectories from both epigenomic and transcriptomic data based on pseudotime estimation.

Single-cell epigenomics can not only be used to resolve static or dynamic structures in heterogeneous cell populations, but can also be used to determine cell-type-specific epigenetic mechanisms underlying differences in cellular activation or dysfunction. In complex scenarios such as in vitro stimulation experiments [47] or prospective clinical disease studies, comparing single-cell epigenome profiles across conditions is thought to expose alterations underlying pathological phenotypes. To circumvent the statistical difficulties arising from the sparse signal of individual genomic regions, chromVAR [52] estimates gain or loss of accessibility across groups of peaks sharing the same transcription factor binding motif or annotation and thereby enables the identification of potentially targetable signaling pathways.

With the advancement of diverse single-cell transcriptomic and epigenomic technologies and the consequential development of single-cell multiomics protocols, such as scNMT-seq [30], computational solutions are required that enable the integration of the different information layers of such data sets to simultaneously and comprehensively study the combined genotypic, regulatory, and phenotypic mechanisms evident only at single-cell resolution. As these technical possibilities emerged only very recently, computational approaches for their analysis are at a very early stage but promise high potential. Although originally developed for bulk multiomics data sets, Multi-Omics Factor Analysis (MOFA) [113] can be applied to single-cell multiomics data sets in order to identify latent factors specific for or shared across data types representing the underlying principal axes of heterogeneity in an unsupervised fashion. Alternatively, the popular single-cell genomics analysis tool kit Seurat [114] enables the integration of chromatin accessibility and gene expression to cross-validate and detect shared and specific regulatory mechanisms.

Studying the immune system using single-cell epigenomics

Historically, in immunological research, flow cytometry is considered as the classic single-cell method, and it is one of the most common techniques to study the immune system. In the context of epigenetics, measuring cell-surface protein levels can be used, for example, to validate the phenotypic output of epigenomic changes, such as the induction of Foxp3 expression in CD8+ T cells upon drug-induced DNA hypomethylation [115]. Additionally, this technique can determine global changes in the epigenome by means of intracellular labeling of methylated DNA [116] or histone marks [117]. As an example, T cell activation can be detected using flow cytometry to profile global changes in chromatin decondensation [118]. Furthermore, epigenetic landscape profiling using cytometry by time-of-flight (EpiTOF) was used to simultaneously detect up to 40 chromatin marks by mass tag-labeled antibodies covering different classes of histone modifications and histone variants [119]. This technique was successfully applied to unveil lineage and cell-type-specific epigenetic profiles of distinct immune cells, building an epigenetic atlas of 22 cell subtypes of the healthy immune system. Moreover, it was demonstrated that interindividual as well as intercellular heterogeneity of chromatin modifications increases throughout the aging process. EpiTOF analysis of monozygotic and dizygotic twins showed that the greater variance in the chromatin modification landscape is caused by nonheritable influences [120].

The simultaneous development of diverse single-cell technologies naturally resulted in their combination in multiomics approaches, with the combination of flow cytometry and downstream single-cell transcriptomics being the most obvious. While a priori established protein markers can be used to preselect and characterize cell populations prior to downstream (multi-)OMICs applications, single-cell RNA-seq can help to readily identify transcripts encoding for novel cell (sub)type-specific marker proteins [121]. In turn, this information can then be used to develop new antibodies for flow cytometry or cell tagging strategies [122]. Following this line of thought, single-cell methods and their combinations clearly offer unprecedented possibilities to identify and thoroughly characterize cellular subpopulations in a targeted or non-targeted way [123, 124]. Both transcriptomic and epigenomic fingerprints allow for the identification of cell heterogeneity of complete tissues or preselected populations [40, 125, 126]. Although proven to be possible, e.g., by scNMT-seq [30], obtaining both epigenetic and transcriptomic information from single cells at scale is still technically challenging. Therefore, several multiomics studies circumvent this issue by combining bulk epigenetic methods with single-cell RNA-seq to mutually support their findings. However, comprehensive studies on the epigenome and the transcriptome of the same cell are imperative to gain insights into the dynamic mechanisms that cannot be captured in bulk. In this section, we present selected studies applying modern NGS-based single-cell epigenomics technologies to explore regulatory mechanisms in the immune system. We focus on three major features of the immune system, namely cellular differentiation during hematopoiesis, malignant transformation in leukemia, and functional deterioration during aging.

Hematopoiesis

Hematopoiesis describes the coordinated differentiation of multipotent stem cells and precursor cells into lineage-restricted mature blood cells through asymmetric cell division [127]. Hence, it serves as an ideal model for exploring the nature of multipotent cell fate decisions and presents a paradigmatic example to study both transcriptomic and epigenomic technologies at single-cell resolution. Combined analysis of single-cell transcriptomic profiles of index-sorted murine bone marrow cells and genome-wide histone modifications in corresponding purified populations has helped to formulate a new model of adult murine myelopoiesis, suggesting early commitment of progenitor cells toward distinct lineages [69, 128]. Correspondingly, droplet-based single-cell Chip-seq was evaluated on a mixture of embryonic stem (ES) cells, fibroblasts, and hematopoietic progenitors and identified subpopulations based on distinct chromatin profiles of pluripotency and differentiation priming. These findings were corroborated by orthogonal single-cell gene expression data revealing that not all aspects of heterogeneity are captured by transcriptional analysis alone and require additional epigenetic information.

In an elegant multiomics study from 2016, Yu et al. linked inter-clonal functional heterogeneity of murine hematopoietic stem cell (HSC) clones to distinct epigenetic and transcriptional patterns by combining endogenous in vivo fluorescent cell tagging, RNA-Seq, Whole Genome Bisulfite DNA Sequencing (WGBS), and ATAC-seq [129]. These findings impressively demonstrated epigenetically driven cell autonomy on a clonal level, which was shown to be stable even under diverse perturbations. Additionally, the authors profiled the transcriptomes of single HSCs from individual clones to resolve intra-clonal variability, reporting that individual clones showed different extents of intercellular heterogeneity. Translating this clonal analysis approach to humans is very difficult as genetic models are not feasible for human research. However, recent proof-of-principle studies demonstrated the potential of lineage tracing analyses in humans based on somatic mutations in the mitochondrial DNA as detected by single-cell ATAC and RNA-seq technologies. Prospectively, these reports foreshadow the immense potential of single-cell omics to fully elucidate both clonal relationships and single-cell heterogeneity during differentiation processes, such as hematopoiesis [130, 131]. Also in 2016, Corces and Buenrostro et al. defined chromatin accessibility and transcription profiles for 13 primary blood cell types spanning the hematopoietic hierarchy in humans and reported that enhancer landscapes better reflect cell identity than mRNA levels. They further corroborated these findings in single-cell accessibility profiles from cells highlighting the necessity of single-cell epigenetic profiling to fully disentangle intra-clonal heterogeneity and eventually identify key determinants of cell fate and lineage decision [10]. In a follow-up study, Buenrostro et al. explored differences in single-cell chromatin accessibility profiles across 10 immunophenotypically defined human hematopoietic cell populations and constructed a chromatin accessibility landscape of human hematopoiesis by the pseudo-temporal ordering of cells along continuous developmental trajectories. Moreover, integration of single-cell expression data allowed for an investigation of the complex regulatory dynamics underlying hematopoiesis by association of regulatory element accessibility and transcription factor or target gene expression [31]. To increase the number of cells per experiment and allow a more unbiased analysis of chromatin accessibility across whole tissues without preselection of cell populations, Lareau et al. developed their "droplet single-cell assay for transposase-accessible chromatin using sequencing" (dsc(i)ATAC-seq) and showcased this technique using the example of chromatin accessibility in LPS stimulated human bone marrow, yielding epigenetic profiles for 136,463 cells [47]. In total, this study reported chromatin accessibility profiles for more than 500,000 cells.

Mechanisms of hematopoietic differentiation were further interrogated by integrating genetic fine-mapping for 16 blood cell traits using genome-wide association studies (GWAS) with single-cell chromatin accessibility profiles. This approach explored patterns of developmental regulation to learn about the pleiotropic processes controlling blood cell production. The authors developed a powerful computational method, g-chromVAR, to discriminate closely related and identify trait-relevant cell populations throughout hematopoiesis on the basis of common genetic variants [132].

Leukemia

Disruption of epigenetic processes can lead to altered gene function and cellular malignancies. In this, hematopoiesis is no exception. In acute myeloid leukemia (AML), recurrent genetic mutations in genes involved in DNA methylation, such as *DNMT3A*, *TET2*, and *IDH1/2*, have been shown to influence disease progression [133–135]. Until recently, most studies were based on bulk analyses of epigenetic marks and thus could not fully recapitulate the evolution of the disease. Thanks to the recent development of single-cell omics techniques, the role of epigenetic cell heterogeneity in cancer progression and drug resistance can now be studied with unprecedented resolution holding great potential of clinically translatable findings.

In 2017, Litzenburger et al. investigated single-cell chromatin accessibility and transcription profiles of K562 leukemic cells through data generated in a previous proof-of-feasibility study [22]. They linked GATA transcription factor expression to epigenomic plasticity affecting drug sensitivity and the clonal dynamics of cancer evolution [49]. In detail, the authors identified a cell surface marker covarying with the cellular epigenomic state that enabled isolation of relevant subpopulations and allowed for downstream functional validation of their omics-based findings. By applying single-cell epigenomics to cells isolated from acute myeloid leukemia (AML) patients and healthy donors, Corces et al. described divergent epigenetic profiles of AML cells corresponding to mixed developmental stages, which is not generally observed in hematopoiesis. Furthermore, they implicated *HOX* genes as key factors in preleukemic hematopoietic stem cells. Additionally, sets of regulatory elements were defined to classify hematopoietic cell types and thus allow for deconvolution of population-level epigenetic profiles into proportions of different cell types [10]. Similarly, joint "multiplexed single-cell reduced representation bisulfite" sequencing (MscRRBS) and scRNA-seq profiling of cells isolated from three chronic lymphocytic leukemia (CLL) patients and four healthy controls exposed the decreased coordination of epigenetic and transcriptional regulation results from increased epigenetic heterogeneity, not previously captured by bulk analyses [28]. In a larger follow-up study, integrated bulk and single-cell epigenome and transcriptome profiles of 22 primary CLL and 13 healthy donor B lymphocyte samples were investigated. Compared to B lymphocytes from healthy individuals, genome-wide mapping of histone modifications in CLL samples exposed an enrichment associated with core- and super-enhancer reprogramming in the vicinity of genes involved in CLL [136], such as *BCL2* [137], *LEF1* [138], and *CTLA4* [139], resulting in dysregulated transcriptional output. These findings further underline the impaired cellular coordination due to epigenetic diversification in CLL and emphasize the adaptive features of the malignancy. Another conceptual study used the single-cell triple omics analysis approach scTrio-seq to study the mechanism by which the transcriptome, genome, and DNA methylome regulate each other at single-cell level, showing that changes in the gene copy number were correlated with transcriptional variation without affecting the DNA methylation status [20]. When applied to 25 single cancer cells, derived from a human hepatocellular carcinoma (HCC) tumor tissue, this technique highlighted distinct cellular subpopulations and demonstrated its potential to dissect the complex genomic, epigenomic, and transcriptomic network within a heterogeneous cell population.

Single-cell omics is not limited to characterization of epigenomic plasticity in heterogeneous cell populations in health and disease but can also be used to study specific epigenetic perturbations and evaluate their potential clinical applications. For instance, CRISPR-mediated deletion of selected transcription factors or chromatin modifiers associated with B-cell malignancies and single-cell epigenetic profiling of B lymphoblasts have been successfully used to investigate consequential chromatin reshaping. While depletion of the chromatin modifier *EZH2* led to an increase in the chromatin accessibility of H3K27me3-associated regions, depletion of the transcription factor *SPI1* decreased the chromatin accessibility of SPI1-motif-containing regions and increased the chromatin accessibility of IRF motif regions [140]. This molecular approach of enabling targeted genetic modifications in combination with single-cell epigenetic profiling has an enormous potential to describe regulatory networks under physiological and pathological conditions.

Aging

Human aging is accompanied by physiological changes in many tissues and blood is not an exception with age-related deterioration of the immune system and amplified systemic inflammation being a common occurrence [141]. A longitudinal study based on cellular phenotyping of blood cell subsets using cell mass cytometry, cytokine responsiveness, and gene expression evaluated immune responsiveness with increasing age and found substantial deterioration [142]. Corresponding to the variation in these immune features, aging also affects the epigenetic landscape of immune cells. WGBS obtained from CD4+ cells from a newborn, a 26-year-old individual, and a centenarian showed gradual DNA hypomethylation, particularly evident in CpG-poor promoters and tissue-specific genes. These findings were further supported in PBMCs using a larger number of samples comprising newborns, middle-aged individuals, and nonagenarians [143]. Additionally, a recent study profiling epigenetic and transcriptomic landscapes in PBMCs and purified monocytes, B cells, and T cells of young and elderly donors exposed cell-type specific alterations in chromatin accessibility essentially in CD8+ memory T cells [144].

Nevertheless, epigenetic studies interrogating heterogeneous tissues or cell populations using bulk technologies might miss important physiologic events, which are of particular importance in complex processes such as aging in which fluctuations in cell compositions are known. A single-cell whole-genome sequencing approach in healthy individuals unveiled a significant acquisition of somatic mutations in B lymphocytes during the aging process: a specific alteration difficult to detect in bulk. The number of mutations varied from less than 500 per cell in newborns to more than 3000 in centenarians, which can be linked to the increased incidence of B cell leukemia and the compromised function of B lymphocytes in the elderly. Interestingly, using ATAC-seq data on B lymphocytes revealed an enrichment in the number of mutations overlapping active chromatin regions [145]. Importantly, chromatin modification heterogeneity increases not only interindividually, but also within the immune cell populations, as demonstrated by single-cell epigenome profiling in aging using mass cytometry [120]. Differential analysis of young and older adults unveiled elevated cell-to-cell variability in chromatin modifications and analysis of a twin cohort demonstrated that aging-associated chromatin alterations are predominantly driven by nonheritable influences.

Apart from blood-borne immune cell populations, the epigenome of microglia also changes during aging which has potential implications in cognitive decline and neurodegeneration. For instance, the DNA hypomethylation of the proximal promoter of the *IL1β* gene was associated with increased cytokine production, through an epigenetic process that seems to be mediated by SIRT1 deficiency [146]. A multiomics study conducted in a mouse model of Alzheimer's disease (AD), a common form of dementia in the elderly, led to the discovery of a novel microglia subtype. Based on single-cell transcriptomic profiling, the authors identified and described a new AD-associated phagocytic cell type, termed Alzheimer's disease-associated microglia (DAM), and defined CD11c as a specific protein marker for cell sorting. Subsequent epigenetic profiling of these cells and homeostatic microglia in bulk using iChIP, however, revealed that the disease-associated transcriptional program is already primed in homeostasis [121]. In 2017, we investigated the effect of intrinsic and extrinsic factors, such as sexual identity and the microbiome, on microglial properties over time and found that the absence of the microbiome in germ-free mice had a time dependent and sexually dimorphic impact both prenatally and postnatally, i.e., microglia were more profoundly perturbed in male embryos and female adults. Single-cell epigenomic technologies in combination with lineage-tracing methods will help to further elucidate the regulatory mechanisms behind these sex-specific differences and explain how extrinsic and intrinsic factors are epigenetically memorized during development.

In contrast to most studies so far using single-cell epigenomics either to prove the technical feasibility of a new protocol or to address a very specific biological phenomenon, a recent publication reported on the successful production of a comprehensive single-cell chromatin accessibility atlas encompassing ∼100,000 cells from 13 adult mouse tissues [33]. Sci-ATAC-seq was used to detect roughly 400,000 differentially accessible genomic elements contributing to 85 distinct patterns of chromatin accessibility. Among a multitude of other interesting findings, the authors observed cell-type specific enrichments of heritability signals for hundreds of complex traits including autoimmune diseases and AD when intersecting mouse chromatin accessibility with human genome-wide association summary statistics. Correspondingly, interrogating the single-cell chromatin accessibility around SNPs from genome-wide association studies (GWAS) in human adult brain cells exposed microglia as the cell type with the most significant enrichment of accessible DNA regions around AD risk variants [29].

Conclusions and future perspectives

Single-cell epigenomics with a spatial resolution

Recently, there have been significant advances in the field of single-cell transcriptomics [147–149], whereby the inclusion of a spatial dimension is opening up our understanding of tissue contexts in transcriptomics. This is the first time we can consider the development of cellular maps and atlases, much like it is standard procedure in geography or radio astronomy.

One could argue that epigenomics has long considered spatial information insofar as the intricacies of three-dimensional folding and localization of DNA within intracellular domains (e.g., topologically, lamina and nucleolar associating domains). However, current single-cell epigenomics methods [20, 29] do not yet enable tissue-wide spatial resolution. Following the rapid advances in both multiomics techniques as well as spatial transcriptomics, we expect the development of single-cell techniques integrating multiple components of the epigenome as well as the transcriptome into combined measurements together with spatial resolution. The isolation of cells via LCM presents one possibility for collecting intercellular spatial data alongside epigenomic information. Whilst it does not allow cells to remain in situ during analysis like advanced spatial transcriptomic methods, it does offer a visual way of tracking cell position before epigenomic profiling of selected cells. Developmental processes, such as organogenesis or adult stem cell differentiation, e.g., gastrointestinal epithelial renewal as well as pathogenic processes, including carcinogenesis, are obvious examples for the vital necessity of such techniques. Understanding how the epigenome is (re-)programmed over time and in space in physiological and pathological conditions is of tremendous importance to identify the roots of complex diseases and determine possibilities for therapeutic intervention.

Other future applications

The number of single-cell epigenomic technologies available has increased greatly in recent years. With this expansion of possible applications and techniques, it is clear that more accurate approaches, both experimentally and computationally, are needed to improve the complexity and precision of chromatin accessibility models. Establishing high-resolution single-cell genome-wide chromatin accessibility, chromatin conformation, histone modification maps, and methylomes may help to elucidate the relationships between the physical states and the functionality of chromatin that would better inform our understanding of disease pathways and therapeutic targets. Of course, this is most efficiently achieved with further developed assays allowing very detailed yet high-throughput data generation.

A current example, scATAC-seq [38] identifying cell-type-specific regulatory regions, is discussed in this review. Of particular interest is the high-throughput droplet-based version dscATAC-seq [47]. Insights derived from such technologies could lead to key developments in the search for therapeutic targets where understanding the specificity in regulatory regions across immune cell populations could prove essential. This could be particularly important in clinical studies where previously bulk methods have required a rather large number of cells to yield sufficient material for robust analyses. This is not the case with single-cell methodologies like scATAC-seq, and so it is now possible to work on a wider range of conditions in humans using relatively small biopsy sample sizes. The benefits of these single-cell technologies are already observable in areas such as oncology, as highlighted for AML [10]. The ability to define crucial points in gene differentiation and mutation on a single-cell level has huge potential in this field, especially if it can be translated into diagnostic and prognostic technologies.

Likewise, we expect novel methods to better address the sparsity of epigenomic single-cell data by seeking to increase method sensitivity both experimentally and computationally or by specifically collecting targeted single-cell information, as it already occurs in single-cell transcriptomics [122]. We are convinced that the heterogeneous nature of cell populations as they are captured via multiomics approaches represents a more detailed approach of depicting novel cellular regulatory networks.

An important aspect of epigenetic studies is the understanding of chromatin function in order to then understand how this can be pharmacologically harnessed. One recent example of a technology that is building toward this goal is CRISPR-Cas9-based epigenomic regulatory element screening (CERES) [150]. This technique uses CRISPR-Cas9 to target noncoding regulatory elements surrounding genes of interest and perform both loss- and gain-of-function screens. This provides a high-throughput method for accurately targeting promoter and enhancer regions in order to comprehend their full range of functionality. These approaches will be central to establish causal relationships between epigenetic regulation and cellular function in future studies.

References

[1] Esteller M. Epigenetics in cancer. New Engl J Med 2008;358(11):1148–59. https://doi.org/10.1056/NEJMra072067.

[2] Arney KL, Fisher AG. Epigenetic aspects of differentiation. J Cell Sci 2004;117(19):4355–63. https://doi.org/10.1242/jcs.01390.

[3] Palini S, et al. Epigenetic regulatory mechanisms during preimplantation embryo development. Ann N Y Acad Sci 2011;1221(1):54–60. https://doi.org/10.1111/j.1749-6632.2010.05937.x.

[4] Ahmed F. Epigenetics: tales of adversity. Nature 2010;468(7327):S20. https://doi.org/10.1038/468S20a.

[5] Fraga MF, et al. Epigenetic differences arise during the lifetime of monozygotic twins. Proc Natl Acad Sci 2005;102(30):10604–9. https://doi.org/10.1073/pnas.0500398102.

[6] Goldberg AD, Allis CD, Bernstein E. Epigenetics: a landscape takes shape. Cell 2007;128(4):635–8. https://doi.org/10.1016/j.cell.2007.02.006.

[7] Lau CM, et al. Epigenetic control of innate and adaptive immune memory. Nat Immunol 2018;19(9):963–72. https://doi.org/10.1038/s41590-018-0176-1.

[8] Moss J, et al. Comprehensive human cell-type methylation atlas reveals origins of circulating cell-free DNA in health and disease. Nat Commun 2018;9(1):5068. https://doi.org/10.1038/s41467-018-07466-6.

[9] Moran S, et al. Epigenetic profiling to classify cancer of unknown primary: a multicentre, retrospective analysis. Lancet Oncol 2016;17(10):1386–95. https://doi.org/10.1016/S1470-2045(16)30297-2.

[10] Corces MR, et al. Lineage-specific and single-cell chromatin accessibility charts human hematopoiesis and leukemia evolution. Nat Genet 2016;48(10):1193–203. https://doi.org/10.1038/ng.3646.

[11] Kundaje A, et al. Integrative analysis of 111 reference human epigenomes. Nature 2015;518(7539):317–30. https://doi.org/10.1038/nature14248.

[12] Lorthongpanich C, et al. Single-cell DNA-methylation analysis reveals epigenetic chimerism in preimplantation embryos. Science 2013;341(6150):1110–2. https://doi.org/10.1126/science.1240617.

[13] Nagano T, et al. Single-cell Hi-C reveals cell-to-cell variability in chromosome structure. Nature 2013;502(7469):59–64. https://doi.org/10.1038/nature12593.

[14] Guo H, et al. Single-cell methylome landscapes of mouse embryonic stem cells and early embryos analyzed using reduced representation bisulfite sequencing. Genome Res 2013;23(12):2126. https://doi.org/10.1101/GR.161679.113.

[15] Smallwood SA, et al. Single-cell genome-wide bisulfite sequencing for assessing epigenetic heterogeneity. Nat Methods 2014;11(8):817–20. https://doi.org/10.1038/nmeth.3035.

[16] Cusanovich DA, et al. Multiplex single-cell profiling of chromatin accessibility by combinatorial cellular indexing. Science 2015;348(6237):910–4. https://doi.org/10.1126/science.aab1601.

[17] Rotem A, et al. Single-cell ChIP-seq reveals cell subpopulations defined by chromatin state. Nat Biotechnol 2015;33(11):1165–72. https://doi.org/10.1038/nbt.3383.

[18] Nagano T, et al. Single-cell Hi-C for genome-wide detection of chromatin interactions that occur simultaneously in a single cell. Nat Protoc 2015;10(12):1986–2003. https://doi.org/10.1038/nprot.2015.127.

[19] Angermueller C, et al. Parallel single-cell sequencing links transcriptional and epigenetic heterogeneity. Nat Methods 2016;13(3):229–32. https://doi.org/10.1038/nmeth.3728.

[20] Hou Y, et al. Single-cell triple omics sequencing reveals genetic, epigenetic, and transcriptomic heterogeneity in hepatocellular carcinomas. Cell Res 2016;26(3):304–19. https://doi.org/10.1038/cr.2016.23.

[21] Hu Y, et al. Simultaneous profiling of transcriptome and DNA methylome from a single cell. Genome Biol 2016;17(1):88. https://doi.org/10.1186/s13059-016-0950-z.

[22] Buenrostro JD, et al. Single-cell chromatin accessibility reveals principles of regulatory variation. Nature 2015;523(7561):486–90. https://doi.org/10.1038/nature14590.

[23] Mooijman D, et al. Single-cell 5hmC sequencing reveals chromosome-wide cell-to-cell variability and enables lineage reconstruction. Nat Biotechnol 2016;34(8):852–6. https://doi.org/10.1038/nbt.3598.

[24] Ramani V, et al. Massively multiplex single-cell Hi-C. Nat Methods 2017;14(3):263–6. https://doi.org/10.1038/nmeth.4155.

[25] Flyamer IM, et al. Single-nucleus Hi-C reveals unique chromatin reorganization at oocyte-to-zygote transition. Nature 2017;544(7648):110–4. https://doi.org/10.1038/nature21711.

[26] Pott S. Simultaneous measurement of chromatin accessibility, DNA methylation, and nucleosome phasing in single cells. elife 2017;6. https://doi.org/10.7554/eLife.23203.

[27] Luo C, et al. Single-cell methylomes identify neuronal subtypes and regulatory elements in mammalian cortex. Science (New York, NY) 2017;357(6351):600–4. https://doi.org/10.1126/science.aan3351.

[28] Chaligne R, et al. Single-cell joint methylomics and transcriptomics define the epigenetic evolution and lineage histories of chronic lymphocytic leukemia. Blood 2017;130(Suppl 1). Available at: http://www.bloodjournal.org/content/130/Suppl_1/55?sso-checked=true. (Accessed 28 June 2019).
[29] Lake BB, et al. Integrative single-cell analysis of transcriptional and epigenetic states in the human adult brain. Nat Biotechnol 2018;36(1):70–80. https://doi.org/10.1038/nbt.4038.
[30] Clark SJ, et al. scNMT-seq enables joint profiling of chromatin accessibility DNA methylation and transcription in single cells. Nat Commun 2018;9(1):781. https://doi.org/10.1038/s41467-018-03149-4.
[31] Buenrostro JD, et al. Integrated single-cell analysis maps the continuous regulatory landscape of human hematopoietic differentiation. Cell 2018;173(6):1535–1548.e16. https://doi.org/10.1016/j.cell.2018.03.074.
[32] Mulqueen RM, et al. Highly scalable generation of DNA methylation profiles in single cells. Nat Biotechnol 2018;36(5):428–31. https://doi.org/10.1038/nbt.4112.
[33] Cusanovich DA, Hill AJ, et al. A single-cell atlas of in vivo mammalian chromatin accessibility. Cell 2018;174(5):1309–1324.e18. https://doi.org/10.1016/j.cell.2018.06.052.
[34] Mezger A, et al. High-throughput chromatin accessibility profiling at single-cell resolution. Nat Commun 2018;9(1):3647. https://doi.org/10.1038/s41467-018-05887-x.
[35] Luo C, et al. Robust single-cell DNA methylome profiling with snmC-seq2. Nat Commun 2018; 9(1):3824. https://doi.org/10.1038/s41467-018-06355-2.
[36] Cao J, et al. Joint profiling of chromatin accessibility and gene expression in thousands of single cells. Science (New York, NY) 2018;361(6409):1380–5. https://doi.org/10.1126/science.aau0730.
[37] Chen X, et al. Joint single-cell DNA accessibility and protein epitope profiling reveals environmental regulation of epigenomic heterogeneity. Nat Commun 2018;9(1):4590. https://doi.org/10.1038/s41467-018-07115-y.
[38] Chen X, et al. A rapid and robust method for single cell chromatin accessibility profiling. Nat Commun 2018;9(1):5345. https://doi.org/10.1038/s41467-018-07771-0.
[39] Lee D-S, et al. Single-cell multi-omic profiling of chromatin conformation and DNA methylome. bioRxiv 2018;503235. https://doi.org/10.1101/503235.
[40] Gu C, et al. Integrative single-cell analysis of transcriptome, DNA methylome and chromatin accessibility in mouse oocytes. Cell Res 2019;29(2):110–23. https://doi.org/10.1038/s41422-018-0125-4.
[41] Liu L, et al. Deconvolution of single-cell multi-omics layers reveals regulatory heterogeneity. Nat Commun 2019;10(1):470. https://doi.org/10.1038/s41467-018-08205-7.
[42] Linker SM, et al. Combined single-cell profiling of expression and DNA methylation reveals splicing regulation and heterogeneity. Genome Biol 2019;20(1):30. https://doi.org/10.1186/s13059-019-1644-0.
[43] Ku WL, et al. Single-cell chromatin immunocleavage sequencing (scChIC-seq) to profile histone modification. Nat Methods 2019;16(4):323–5. https://doi.org/10.1038/s41592-019-0361-7.
[44] Sinnamon JR, et al. The accessible chromatin landscape of the murine hippocampus at single-cell resolution. Genome Res 2019;29(5):857–69. https://doi.org/10.1101/gr.243725.118.
[45] Satpathy AT, et al. Massively parallel single-cell chromatin landscapes of human immune cell development and intratumoral T cell exhaustion. bioRxiv 2019;610550. https://doi.org/10.1101/610550.
[46] Kaya-Okur HS, et al. CUT&Tag for efficient epigenomic profiling of small samples and single cells. Nat Commun 2019;10(1):1930. https://doi.org/10.1038/s41467-019-09982-5.
[47] Lareau CA, et al. Droplet-based combinatorial indexing for massive-scale single-cell chromatin accessibility. Nat Biotechnol 2019;1. https://doi.org/10.1038/s41587-019-0147-6.
[48] Chung C-Y, et al. Single-cell chromatin accessibility analysis of mammary gland development reveals cell state transcriptional regulators and cellular lineage relationships. bioRxiv 2019;624957. https://doi.org/10.1101/624957.
[49] Litzenburger UM, et al. Single-cell epigenomic variability reveals functional cancer heterogeneity. Genome Biol 2017;18(1):15. https://doi.org/10.1186/s13059-016-1133-7.
[50] Bravo González-Blas C, et al. cisTopic: cis-regulatory topic modeling on single-cell ATAC-seq data. Nat Methods 2019;16(5):397–400. https://doi.org/10.1038/s41592-019-0367-1.
[51] Pliner HA, et al. Cicero predicts cis-regulatory DNA interactions from single-cell chromatin accessibility data. Mol Cell 2018;71(5):858–871.e8. https://doi.org/10.1016/J.MOLCEL.2018.06.044.

[52] Schep AN, et al. chromVAR: inferring transcription-factor-associated accessibility from single-cell epigenomic data. Nat Methods 2017;14(10):975–8. https://doi.org/10.1038/nmeth.4401.

[53] Lavin Y, et al. Tissue-resident macrophage enhancer landscapes are shaped by the local microenvironment. Cell 2014;159(6):1312–26. https://doi.org/10.1016/j.cell.2014.11.018.

[54] Farlik M, et al. Single-cell DNA methylome sequencing and bioinformatic inference of epigenomic cell-state dynamics. Cell Rep 2015;10(8):1386–97. https://doi.org/10.1016/j.celrep.2015.02.001.

[55] Schutsky EK, et al. Nondestructive, base-resolution sequencing of 5-hydroxymethylcytosine using a DNA deaminase. Nat Biotechnol 2018;36(11):1083–90. https://doi.org/10.1038/nbt.4204.

[56] Kind J, et al. Genome-wide maps of nuclear lamina interactions in single human cells. Cell 2015;163(1): 134–47. https://doi.org/10.1016/j.cell.2015.08.040.

[57] Wu F, Olson BG, Yao J. DamID-seq: genome-wide mapping of protein-DNA interactions by high throughput sequencing of adenine-methylated DNA fragments. J Vis Exp 2016;107:e53620. https://doi.org/10.3791/53620.

[58] Schones DE, et al. Dynamic regulation of nucleosome positioning in the human genome. Cell 2008;132(5):887–98. https://doi.org/10.1016/j.cell.2008.02.022.

[59] Kelly TK, et al. Genome-wide mapping of nucleosome positioning and DNA methylation within individual DNA molecules. Genome Res 2012;22(12):2497–506. https://doi.org/10.1101/gr.143008.112.

[60] Jin W, et al. Genome-wide detection of DNase I hypersensitive sites in single cells and FFPE tissue samples. Nature 2015;528(7580):142–6. https://doi.org/10.1038/nature15740.

[61] Mulqueen RM, et al. Improved single-cell ATAC-seq reveals chromatin dynamics of in vitro corticogenesis. bioRxiv 2019;637256. https://doi.org/10.1101/637256.

[62] Sos BC, et al. Characterization of chromatin accessibility with a transposome hypersensitive sites sequencing (THS-seq) assay. Genome Biol 2016;17(1):20. https://doi.org/10.1186/s13059-016-0882-7.

[63] Belton J-M, et al. Hi–C: a comprehensive technique to capture the conformation of genomes. Methods 2012;58(3):268–76. https://doi.org/10.1016/j.ymeth.2012.05.001.

[64] Duan Z, et al. A genome-wide 3C-method for characterizing the three-dimensional architectures of genomes. Methods 2012;58(3):277–88. https://doi.org/10.1016/j.ymeth.2012.06.018.

[65] Picelli S. Single-cell RNA-sequencing: the future of genome biology is now. RNA Biol 2017;14(5): 637–50. https://doi.org/10.1080/15476286.2016.1201618.

[66] Hu P, et al. Single cell isolation and analysis. In: Frontiers in cell and developmental biology, vol. 4. Frontiers Media SA; 2016. p. 116. https://doi.org/10.3389/fcell.2016.00116.

[67] Carlos AF, et al. From brain collections to modern brain banks: a historical perspective. Alzheimers Dement (NY) 2019;5:52–60. https://doi.org/10.1016/j.trci.2018.12.002.

[68] Gross A, et al. Technologies for single-cell isolation. Int J Mol Sci 2015;16(8):16897. https://doi.org/10.3390/IJMS160816897.

[69] Paul F, et al. Transcriptional heterogeneity and lineage commitment in myeloid progenitors. Cell 2015;163(7):1663–77. https://doi.org/10.1016/j.cell.2015.11.013.

[70] Herman J, et al. High-throughput single cell sequencing using CEL-Seq2 on a nanoliter dispensing robot. Mol Biol 2018. https://doi.org/10.1038/protex.2018.069.

[71] Macosko EZ, et al. Highly parallel genome-wide expression profiling of individual cells using nanoliter droplets. Cell 2015;161(5):1202–14. https://doi.org/10.1016/j.cell.2015.05.002.

[72] Durruthy-Durruthy R, Ray M. Using fluidigm C1 to generate single-cell full-length cDNA libraries for mRNA sequencing. New York, NY: Humana Press; 2018. p.199–221. https://doi.org/10.1007/978-1-4939-7471-9_11.

[73] Zheng GXY, et al. Massively parallel digital transcriptional profiling of single cells. Nat Commun 2017;8(1):14049. https://doi.org/10.1038/ncomms14049.

[74] Gierahn TM, et al. Seq-well: portable, low-cost RNA sequencing of single cells at high throughput. Nat Methods 2017;14(4):395–8. https://doi.org/10.1038/nmeth.4179.

[75] Nichterwitz S, et al. LCM-seq: a method for spatial transcriptomic profiling using laser capture microdissection coupled with PolyA-based RNA sequencing. Methods Mol Biol 2018;95–110. https://doi.org/10.1007/978-1-4939-7213-5_6.

[76] Schillebeeckx M, et al. Laser capture microdissection-reduced representation bisulfite sequencing (LCM-RRBS) maps changes in DNA methylation associated with gonadectomy-induced

adrenocortical neoplasia in the mouse. Nucleic Acids Res 2013;41(11):e116. https://doi.org/10.1093/nar/gkt230.
[77] Ramani V, et al. Sci-Hi-C: a single-cell Hi-C method for mapping 3D genome organization in large number of single cells. bioRxiv 2019;579573. https://doi.org/10.1101/579573.
[78] Nawy T. Single-cell sequencing. Nat Methods 2014;11(1):18. https://doi.org/10.1038/nmeth.2771.
[79] Gaiti F, et al. Epigenetic evolution and lineage histories of chronic lymphocytic leukaemia. Nature 2019;569(7757):576–80. https://doi.org/10.1038/s41586-019-1198-z.
[80] Miura F, et al. Amplification-free whole-genome bisulfite sequencing by post-bisulfite adaptor tagging. Nucleic Acids Res 2012;40(17):e136. https://doi.org/10.1093/NAR/GKS454.
[81] Booth MJ, et al. Quantitative sequencing of 5-methylcytosine and 5-hydroxymethylcytosine at single-base resolution. Science 2012;336(6083):934–7. https://doi.org/10.1126/science.1220671.
[82] Yu M, et al. Base-resolution analysis of 5-hydroxymethylcytosine in the mammalian genome. Cell 2012;149(6):1368–80. https://doi.org/10.1016/j.cell.2012.04.027.
[83] Sun Z, et al. High-resolution enzymatic mapping of genomic 5-hydroxymethylcytosine in mouse embryonic stem cells. Cell Rep 2013;3(2):567–76. https://doi.org/10.1016/j.celrep.2013.01.001.
[84] Ficz G, et al. Dynamic regulation of 5-hydroxymethylcytosine in mouse ES cells and during differentiation. Nature 2011;473(7347):398–402. https://doi.org/10.1038/nature10008.
[85] Olins DE, Olins AL. Chromatin history: our view from the bridge. Nat Rev Mol Cell Biol 2003; 4(10):809–14. https://doi.org/10.1038/nrm1225.
[86] Song L, Crawford GE. DNase-seq: a high-resolution technique for mapping active gene regulatory elements across the genome from mammalian cells. Cold Spring Harb Protoc 2010;2010(2). p. pdb.prot5384. https://doi.org/10.1101/pdb.prot5384.
[87] Simon JM, et al. Using formaldehyde-assisted isolation of regulatory elements (FAIRE) to isolate active regulatory DNA. Nat Protoc 2012;7(2):256–67. https://doi.org/10.1038/nprot.2011.444.
[88] Buenrostro JD, et al. Transposition of native chromatin for fast and sensitive epigenomic profiling of open chromatin, DNA-binding proteins and nucleosome position. Nat Methods 2013;10(12): 1213–8. https://doi.org/10.1038/nmeth.2688.
[89] Adey A, et al. Rapid, low-input, low-bias construction of shotgun fragment libraries by high-density in vitro transposition. Genome Biol 2010;11(12):R119. https://doi.org/10.1186/gb-2010-11-12-r119.
[90] Liashkovich I, et al. Clathrin inhibitor Pitstop-2 disrupts the nuclear pore complex permeability barrier. Sci Rep 2015;5:9994. https://doi.org/10.1038/srep09994.
[91] Han H, et al. DNA methylation directly silences genes with non-CpG island promoters and establishes a nucleosome occupied promoter. Hum Mol Genet 2011;20(22):4299–310. https://doi.org/10.1093/hmg/ddr356.
[92] Nagano T, et al. Cell-cycle dynamics of chromosomal organization at single-cell resolution. Nature 2017;547(7661):61–7. https://doi.org/10.1038/nature23001.
[93] Macaulay IC, et al. G&T-seq: parallel sequencing of single-cell genomes and transcriptomes. Nat Methods 2015;12(6):519–22. https://doi.org/10.1038/nmeth.3370.
[94] Dey SS, et al. Integrated genome and transcriptome sequencing of the same cell. Nat Biotechnol 2015;33(3):285–9. https://doi.org/10.1038/nbt.3129.
[95] Ji Z, Zhou W, Ji H. Single-cell regulome data analysis by SCRAT. Bioinformatics Edited by I. Birol. 2017;33(18):2930–2. https://doi.org/10.1093/bioinformatics/btx315.
[96] Baker SM, et al. Classifying cells with scasat, a single-cell ATAC-seq analysis tool. Nucleic Acids Res 2019;47(2):e10. https://doi.org/10.1093/nar/gky950.
[97] Danese A, et al. EpiScanpy: integrated single-cell epigenomic analysis. bioRxiv 2019;648097. https://doi.org/10.1101/648097.
[98] Bolger AM, Lohse M, Usadel B. Trimmomatic: a flexible trimmer for Illumina sequence data. Bioinformatics 2014;30(15):2114–20. https://doi.org/10.1093/bioinformatics/btu170.
[99] Langmead B, Salzberg SL. Fast gapped-read alignment with Bowtie 2. Nat Methods 2012;9(4):357–9. https://doi.org/10.1038/nmeth.1923.
[100] Amemiya HM, Kundaje A, Boyle AP. The ENCODE blacklist: identification of problematic regions of the genome. Sci Rep 2019;9(1):9354. https://doi.org/10.1038/s41598-019-45839-z.
[101] Zhang Y, et al. Model-based analysis of ChIP-seq (MACS). Genome Biol 2008;9(9):R137. https://doi.org/10.1186/gb-2008-9-9-r137.

[102] Fang R, et al. Fast and accurate clustering of single cell epigenomes reveals cis-regulatory elements in rare cell types. bioRxiv 2019;615179. https://doi.org/10.1101/615179.

[103] Angermueller C, et al. DeepCpG: accurate prediction of single-cell DNA methylation states using deep learning. Genome Biol 2017;18(1):67. https://doi.org/10.1186/s13059-017-1189-z.

[104] Liu T, Wang Z. scHiCNorm: a software package to eliminate systematic biases in single-cell Hi-C data. Bioinformatics Edited by J. Hancock. 2018;34(6):1046–7. https://doi.org/10.1093/bioinformatics/btx747.

[105] Zhu H, Wang Z. SCL: a lattice-based approach to infer 3D chromosome structures from single-cell Hi-C data. Bioinformatics Edited by A. Valencia. 2019. https://doi.org/10.1093/bioinformatics/btz181.

[106] Blei DM, Lafferty JD, Lafferty JD. Topic models. Chapman and Hall/CRC; 2009. p.101–24. https://doi.org/10.1201/9781420059458-12.

[107] Cusanovich DA, Reddington JP, et al. The cis-regulatory dynamics of embryonic development at single-cell resolution. Nature 2018;555(7697):538–42. https://doi.org/10.1038/nature25981.

[108] Urrutia E, et al. Destin: toolkit for single-cell analysis of chromatin accessibility. Bioinformatics Edited by B. Berger. 2019. https://doi.org/10.1093/bioinformatics/btz141.

[109] van der Maaten L, Hinton G. Visualizing data using t-SNE. J Mach Learn Res 2008;9(Nov):2579–605. Available at: http://www.jmlr.org/papers/v9/vandermaaten08a.html. (Accessed 4 July 2019).

[110] McInnes L, Healy J, Melville J. UMAP: uniform manifold approximation and projection for dimension reduction. Available at: http://arxiv.org/abs/1802.03426; 2018. (Accessed 4 July 2019).

[111] Haghverdi L, et al. Batch effects in single-cell RNA-sequencing data are corrected by matching mutual nearest neighbors. Nat Biotechnol 2018;36(5):421–7. https://doi.org/10.1038/nbt.4091.

[112] Chen H, et al. Single-cell trajectories reconstruction, exploration and mapping of omics data with STREAM. Nat Commun 2019;10(1):1903. https://doi.org/10.1038/s41467-019-09670-4.

[113] Argelaguet R, et al. Multi-Omics Factor Analysis—a framework for unsupervised integration of multi-omics data sets. Mol Syst Biol 2018;14(6):e8124. https://doi.org/10.15252/msb.20178124.

[114] Stuart T, et al. Comprehensive integration of single-cell data. Cell 2019;177(7):1888–1902.e21. https://doi.org/10.1016/J.CELL.2019.05.031.

[115] Freudenberg K, et al. Critical role of TGF-β and IL-2 receptor signaling in Foxp3 induction by an inhibitor of DNA methylation. Front Immunol 2018;9:125. https://doi.org/10.3389/fimmu.2018.00125.

[116] Çelik-Uzuner S, et al. Measurement of global DNA methylation levels by flow cytometry in mouse fibroblasts. In Vitro Cell Dev Biol Anim 2017;53(1):1–6. https://doi.org/10.1007/s11626-016-0075-4.

[117] Watson M, et al. The study of epigenetic mechanisms based on the analysis of histone modification patterns by flow cytoametry. Cytometry A 2014;85(1):78–87. https://doi.org/10.1002/cyto.a.22344.

[118] Bingham KN, Lee MD, Rawlings JS. The use of flow cytometry to assess the state of chromatin in T cells. J Vis Exp 2015;106:e53533. https://doi.org/10.3791/53533.

[119] Cheung P, Vallania F, Dvorak M, et al. Single-cell epigenetics—chromatin modification atlas unveiled by mass cytometry. Clin Immunol 2018;196:40–8. https://doi.org/10.1016/j.clim.2018.06.009.

[120] Cheung P, Vallania F, Warsinske HC, et al. Single-cell chromatin modification profiling reveals increased epigenetic variations with aging. Cell 2018;173(6):1385–1397.e14. https://doi.org/10.1016/J.CELL.2018.03.079.

[121] Keren-Shaul H, et al. A unique microglia type associated with restricting development of Alzheimer's disease. Cell 2017;169(7):1276–1290.e17. https://doi.org/10.1016/j.cell.2017.05.018.

[122] Stoeckius M, et al. Simultaneous epitope and transcriptome measurement in single cells. Nat Methods 2017;14(9):865–8. https://doi.org/10.1038/nmeth.4380.

[123] Nguyen A, et al. Single cell RNA sequencing of rare immune cell populations. Front Immunol 2018;9:1553. https://doi.org/10.3389/fimmu.2018.01553.

[124] Zheng C, et al. Landscape of infiltrating T cells in liver cancer revealed by single-cell sequencing. Cell 2017;169(7):1342–1356.e16. https://doi.org/10.1016/j.cell.2017.05.035.

[125] Delile J, et al. Single cell transcriptomics reveals spatial and temporal dynamics of gene expression in the developing mouse spinal cord. Development 2019;146(12):dev173807. https://doi.org/10.1242/dev.173807.

[126] Satpathy AT, et al. Transcript-indexed ATAC-seq for precision immune profiling. Nat Med 2018; 24(5):580–90. https://doi.org/10.1038/s41591-018-0008-8.

[127] Beckmann J, et al. Asymmetric cell division within the human hematopoietic stem and progenitor cell compartment: identification of asymmetrically segregating proteins. Blood 2007;109(12):5494–501. https://doi.org/10.1182/blood-2006-11-055921.

[128] Schultze JL, Beyer M. Myelopoiesis reloaded: single-cell transcriptomics leads the way. Immunity 2016;44(1):18–20. https://doi.org/10.1016/j.immuni.2015.12.019.

[129] Yu VWC, et al. Epigenetic memory underlies cell-autonomous heterogeneous behavior of hematopoietic stem cells. Cell 2017;168(5):944–5. https://doi.org/10.1016/j.cell.2017.02.010.

[130] Ludwig LS, et al. Lineage tracing in humans enabled by mitochondrial mutations and single-cell genomics. Cell 2019;176(6):1325–1339.e22. https://doi.org/10.1016/J.CELL.2019.01.022.

[131] Xu J, et al. Single-cell lineage tracing by endogenous mutations enriched in transposase accessible mitochondrial DNA. elife 2019;8. https://doi.org/10.7554/eLife.45105.

[132] Ulirsch JC, et al. Interrogation of human hematopoiesis at single-cell and single-variant resolution. Nat Genet 2019;51(4):683–93. https://doi.org/10.1038/s41588-019-0362-6.

[133] Cancer Genome Atlas Research Network, et al. Genomic and epigenomic landscapes of adult de novo acute myeloid leukemia. New Engl J Med 2013;368(22):2059–74. https://doi.org/10.1056/NEJMoa1301689.

[134] Chan SM, Majeti R. Role of DNMT3A, TET2, and IDH1/2 mutations in pre-leukemic stem cells in acute myeloid leukemia. Int J Hematol 2013;98(6):648–57. https://doi.org/10.1007/s12185-013-1407-8.

[135] Ley TJ, et al. *DNMT3A* mutations in acute myeloid leukemia. New Engl J Med 2010;363(25): 2424–33. https://doi.org/10.1056/NEJMoa1005143.

[136] Pastore A, et al. Corrupted coordination of epigenetic modifications leads to diverging chromatin states and transcriptional heterogeneity in CLL. Nat Commun 2019;10(1):1874. https://doi.org/10.1038/s41467-019-09645-5.

[137] Robertson LE, et al. Bcl-2 expression in chronic lymphocytic leukemia and its correlation with the induction of apoptosis and clinical outcome. Leukemia 1996;10(3):456–9. Available at http://www.ncbi.nlm.nih.gov/pubmed/8642861. (Accessed 25 June 2019).

[138] Gutierrez A, et al. LEF-1 is a prosurvival factor in chronic lymphocytic leukemia and is expressed in the preleukemic state of monoclonal B-cell lymphocytosis. Blood 2010;116(16):2975–83. https://doi.org/10.1182/blood-2010-02-269878.

[139] Do P, et al. Leukemic B cell CTLA-4 suppresses costimulation of T cells. J Immunol 2019;202(9): 2806–16. https://doi.org/10.4049/jimmunol.1801359.

[140] Rubin AJ, et al. Coupled single-cell CRISPR screening and epigenomic profiling reveals causal gene regulatory networks'. Cell 2019;176(1–2):361–376.e17. https://doi.org/10.1016/j.cell.2018. 11.022.

[141] Chen H, Zheng X, Zheng Y. Age-associated loss of lamin-B leads to systemic inflammation and gut hyperplasia. Cell 2014;159(4):829–43. https://doi.org/10.1016/j.cell.2014.10.028.

[142] Alpert A, et al. A clinically meaningful metric of immune age derived from high-dimensional longitudinal monitoring. Nat Med 2019;25(3):487–95. https://doi.org/10.1038/s41591-019-0381-y.

[143] Heyn H, et al. Distinct DNA methylomes of newborns and centenarians. Proc Natl Acad Sci 2012; 109(26):10522–7. https://doi.org/10.1073/pnas.1120658109.

[144] Ucar D, et al. The chromatin accessibility signature of human immune aging stems from CD8+ T cells. J Exp Med 2017;214(10):3123–44. https://doi.org/10.1084/jem.20170416.

[145] Zhang L, et al. Single-cell whole-genome sequencing reveals the functional landscape of somatic mutations in B lymphocytes across the human lifespan. Proc Natl Acad Sci 2019;116(18):9014–9. https://doi.org/10.1073/pnas.1902510116.

[146] Cho S-H, et al. SIRT1 deficiency in microglia contributes to cognitive decline in aging and neurodegeneration via epigenetic regulation of IL-1β. J Neurosci 2015;35(2):807–18. https://doi.org/10.1523/JNEUROSCI.2939-14.2015.

[147] Codeluppi S, et al. Spatial organization of the somatosensory cortex revealed by osmFISH. Nat Methods 2018;15(11):932–5. https://doi.org/10.1038/s41592-018-0175-z.

[148] Svensson V, Teichmann SA, Stegle O. SpatialDE: identification of spatially variable genes. Nat Methods 2018;15(5):343–6. https://doi.org/10.1038/nmeth.4636.

[149] Wang X, et al. Three-dimensional intact-tissue sequencing of single-cell transcriptional states. Science (New York, NY) 2018;361(6400):eaat5691. https://doi.org/10.1126/science.aat5691.

[150] Klann TS, et al. CRISPR–Cas9 epigenome editing enables high-throughput screening for functional regulatory elements in the human genome. Nat Biotechnol 2017;35(6):561–8. https://doi.org/10.1038/nbt.3853.

CHAPTER 9

Machine learning and deep learning for the advancement of epigenomics

Magdalena A. Machnicka, Bartek Wilczynski
Institute of Informatics, Faculty of Mathematics, Informatics and Mechanics, University of Warsaw, Poland

Contents

The "epigenetic code" problem 217
Progress of machine learning: Classification versus non-supervised learning 218
Unsupervised approaches 221
Supervised approaches 223
  Methods for training data generation 225
  Classical classification methods 226
  New approaches—Deep learning 231
Conclusion 233
Acknowledgments 234
References 234

The "epigenetic code" problem

In the recent years, as we have been gaining more knowledge about the epigenetic mechanisms and their impact on the state of the cells, it has become clear that given their complexity, we are in need of a system for organizing established facts regarding the epigenetic processes and the way they interact with each other to give rise to very complex epigenetic landscapes of thousands of cell types that can be reproducibly obtained in development [1]. Since there are many intertwined modes of epigenetic regulation including DNA methylation [2], histone modifications [3], and chromosome conformation changes [4], it is important not only to measure and identify each of those components individually, but also to relate them to one another and view them in a systematic way in the context of the cell state and gene expression. So far, despite many individual examples of clear interactions and synergy between these types of epigenetic mechanisms [5–7], we are still far from having a good system even to describe the overall interactions between all of the known types of epigenetic data. Such a system, frequently referred to as the elusive "epigenetic code", is currently very far from our grasp, and for the most part we are unable to even clearly define the exact components of such a model. One thing is clear that if we derived such a global

Epigenetics of the Immune System
https://doi.org/10.1016/B978-0-12-817964-2.00009-5

© 2020 Elsevier Inc.
All rights reserved.

"epigenetic code" even for a relatively simple multicellular species, then it should be able to predict genome-wide changes to a very multidimensional epigenetic state that happen both spontaneously in development and upon external perturbation, be it through stress conditions or genetic modification. We know that we can exert very strong changes on various parts of the epigenetic system, for example by heat shock [8] or by histone deacetylase (HDAC) protein mutants [9], however, we are still very far from good predictive models allowing us to reliably anticipate what epigenetic changes to expect upon perturbation.

Not surprisingly, given the complexity of the system combined with the abundance of high-throughput techniques for the measurement of the epigenetic state of the cells, researchers are now practically forced to use computers even to store and manipulate the epigenomic data. Since the data is already processed by computers, and finding the "epigenetic code" is frequently considered to be the overarching goal of researchers, it is natural to turn to the methodologies developed by computer scientists and statisticians for analyses of big data in other fields. In this chapter, we will cover the most frequently used approaches involving classical machine learning as well as the most recent advances of deep learning approaches in the field of epigenomics.

Progress of machine learning: Classification versus non-supervised learning[a]

During the same time as our knowledge of the epigenetic regulatory system has grown, there has been tremendous development in the field of machine learning (ML). Conceptually, there has been a lot of fundamental ideas advancing artificial intelligence (AI) and artificial neural networks put forth in the 1940s and 1950s of the 20th century, just to mention the development of the artificial neuron models by McCulloch and Pitts [10], Rosenblatt's perceptron [11], and the foundations of general AI laid out by Alan Turing [12]. However, from the point of view of practical application, the real progress in this area needed the booming developments of the computers that really accelerated in the 1980s. Then, with the availability of desktop computers, the dream of computers helping humans in the crucial tasks of decision-making in medicine and business started to gain popularity.

---

[a] While we have tried to summarize the developments in the machine learning field that are important to the topic of this chapter in this section, naturally the scope of a historic perspective on ML is far too wide to be covered here. For the readers who would like to read more on the topic, we recommend the introduction section to the excellent book by Goodfellow, Bengio, and Courvillle [64] for a much deeper text on the matter.

The so-called decision systems were very popular [13] then and were usually represented by computerized versions of decision trees [14]. The science of building optimal decision trees was quickly developing and finding its use in various fields of business and finance; however, its use in medicine was severely hampered by the quickly observed phenomenon that these systems were prone to overfitting, i.e., the propensity of models built on limited training sets of examples is to perform very well on the training examples, but make stupid mistakes on new examples not present in the training data [15]. This was not a problem in technical applications, where it was possible to simply scale up the size of the training set and build commercially successful models especially suited for closed systems such as trading. However, even there, the systems are beating humans mostly with the speed of simple buy/sell decisions and not by some long-term predictive vision. However, it can be safely said that early approaches to the problem of training machines to learn complex concepts were not nearly as successful as it was advertised by the early advocates of machine learning.

While the early approaches to decision-making were based on decision trees and relatively simple neural networks, i.e., both inspired by conscious or unconscious processes involved in human decision-making, they both suffered from the problem of overfitting to relatively small datasets. The solution to these problems, at least partially, lies in the developments coming from the rigorous mathematical description of the generalization problem and the statistical approach to the predictor performance. Crucially, the so-called statistical learning theory [16] posits the problem of learning as a statistical scenario, where we are only faced with a random sample of the training examples drawn from a vast space of possible examples to be decided on. And we consider our task to have a model that performs well on this greater space rather than on the small training sample. While this might sound obvious now, this was far from that commonly accepted in the 1990s when many decision tree-based and similar methods were developed. However, once the statistical approach to the learning problem was proposed, it led to quick adoption of both randomized strategies for model building (such as very successful Random Forest method by Breiman [17]) as well as randomized strategies for model testing, most prominently represented by various bootstrapping strategies and cross-validation methods [18]. All these developments in the late 1990s allowed the growing machine learning field to consolidate and develop strong tools for the assessment of model quality as well as problem difficulty. It allowed researchers to widely accept the notion that a model that performs worse on the training data might be better in generalizing over new examples and that we should focus on the generalization in learning instead of training set performance optimization [15].

In these years, in parallel with the decision support, or classification system development, the machine learning community was also developing techniques for the so-called unsupervised learning. While in the previously discussed field of supervised learning, we are given a training set of examples where the correct decision is given and our goal is to build a system that is able to make such decisions both in the training examples and in new, unseen cases, the unsupervised approach is more focused on finding similarities and differences between the samples in the training set. If we formalize the decision-making problem as a classification task, where any given sample is equipped with a set of attributes, the task for the model is to predict which of the finite set of classes the sample came from, based on the values of the attributes. In the most standard case, we can have a set of patients with a typical set of clinical measurements attached and the decision-making system would need to either classify them as healthy or diseased (binary classification) or classify the patients into multiple subtypes of the same disease (very frequent in case of various tumors with subtypes responding to different treatments [19]). While in supervised approach the typical training set would consist of the set of patients with both measurements and the decision (e.g., their known disease subtype as in Ref. [20]) in the unsupervised learning scenario, we would be faced with just the set of patients without the classes, but we would expect the machine learning method to identify the set of sample subtypes purely based on the observed attributes and their similarity. Typically, this is solved by clustering methods that achieve the goal of class identification either by grouping the samples into a predefined number of classes such as in the k-means and related methods or by hierarchically aggregating most similar samples (agglomerative clustering, such as in Ref. [21]) or dividing the most dissimilar groups (divisive clustering, in Ref. [22]) and creating the hierarchy of clusters. Many of these algorithms are related to the problem of dimensionality reduction, performed by methods such as principal components analysis (PCA) or multidimensional scaling (MDS) (see Ref. [16] for a review) and in case of missing data there are other methods employing various kinds of expectation–maximization techniques, such as Hidden Markov Models.

In recent years, we have observed the comeback of the neural network models to the front of the field of machine learning. This is in part connected to the technical developments allowing for the usage of highly specialized hardware, such as field-programmable gate arrays (FPGAs) [23] and graphics processing units (GPUs) [24], for very fast computations, as well as the developments of the deep architectures for neural networks with many layers [25] and finally mixed models with some of the layers of the networks being traditionally fully connected, while some others having more regular connection topologies such as convolutional networks [26].

All of the mentioned methodologies have found some application in the field of epigenomic data analysis. Since this is the most natural distinction, we will discuss the unsupervised methods first and then move to the supervised approaches. Within the supervised approaches, we will first cover examples of the application of the more classical ML methodologies including Bayesian Networks [27], Random Forests [17], and Support Vector Machines [28]. Then, in the last section, we will cover the newest approaches based on the deep learning paradigm, which is the most promising approach for the foreseeable future.

Unsupervised approaches

Unsupervised machine learning approaches are often applied to the analysis of complex signals from different types of high-throughput genome-wide experiments describing chromatin states. The standard and probably the most common method is ChromHMM [29, 30] which uses Hidden Markov Model (HMM) to perform segmentation of chromatin based on signal from diverse experiment types. This methodology can be seen as a type of dimensionality reduction method. By providing an argument defining the desired number of output states the user can transform multidimensional dataset of genome-wide signals from different chromatin marks into low-dimensionality dataset, which is much easier to interpret. ChromHMM was applied to the ENCODE dataset comprising results of open chromatin assays DNase-seq and FAIRE and ChIP-seq on eight histone modifications, RNA polymerase 2 (Pol2), and CTCF, performed on six different cell lines to segment human genome into 25 states [31] (Fig. 1).

A similar analysis was performed several years earlier by Filion et al. who have defined five principal chromatin types in *Drosophila* cells by combining principal component analysis and HMM [32]. Values from the first three principal components of the dataset comprising genome-wide location maps of 53 chromatin proteins and four histone modifications were used to perform segmentation by five-state HMM.

It must be kept in mind, however, that the number of states identified by this category of methods is arbitrarily chosen and provided as an input argument. Thus, biological interpretation of the segmentation results must be performed post hoc, using expert knowledge about expected biological consequences of the presence of particular combination of chromatin marks. Additionally, the arbitrary assignment of the number of states increases the risk that the model will not distinguish between states having relatively similar characteristics but highly different function (e.g., enhancers vs promoters or active vs inactive enhancers). Discovery of rare but functionally important and distinct combination of chromatin marks is also difficult within this unsupervised framework. Overall, segmentation methods based on unsupervised machine learning allow to greatly increase the resolution of chromatin analysis from simple, binary division to euchromatin and heterochromatin but it should be kept in mind that they can be applied to identify trends, not to perform primary functional analysis.

Among other unsupervised approaches used in the analysis of epigenomic data, Segway [33] is a method which performs genome segmentation based on signals from epigenomic marks, very similar to ChromHMM but using a dynamic Bayesian network approach. Zinba [34] is an algorithm for enrichment identification in ChIP-seq,

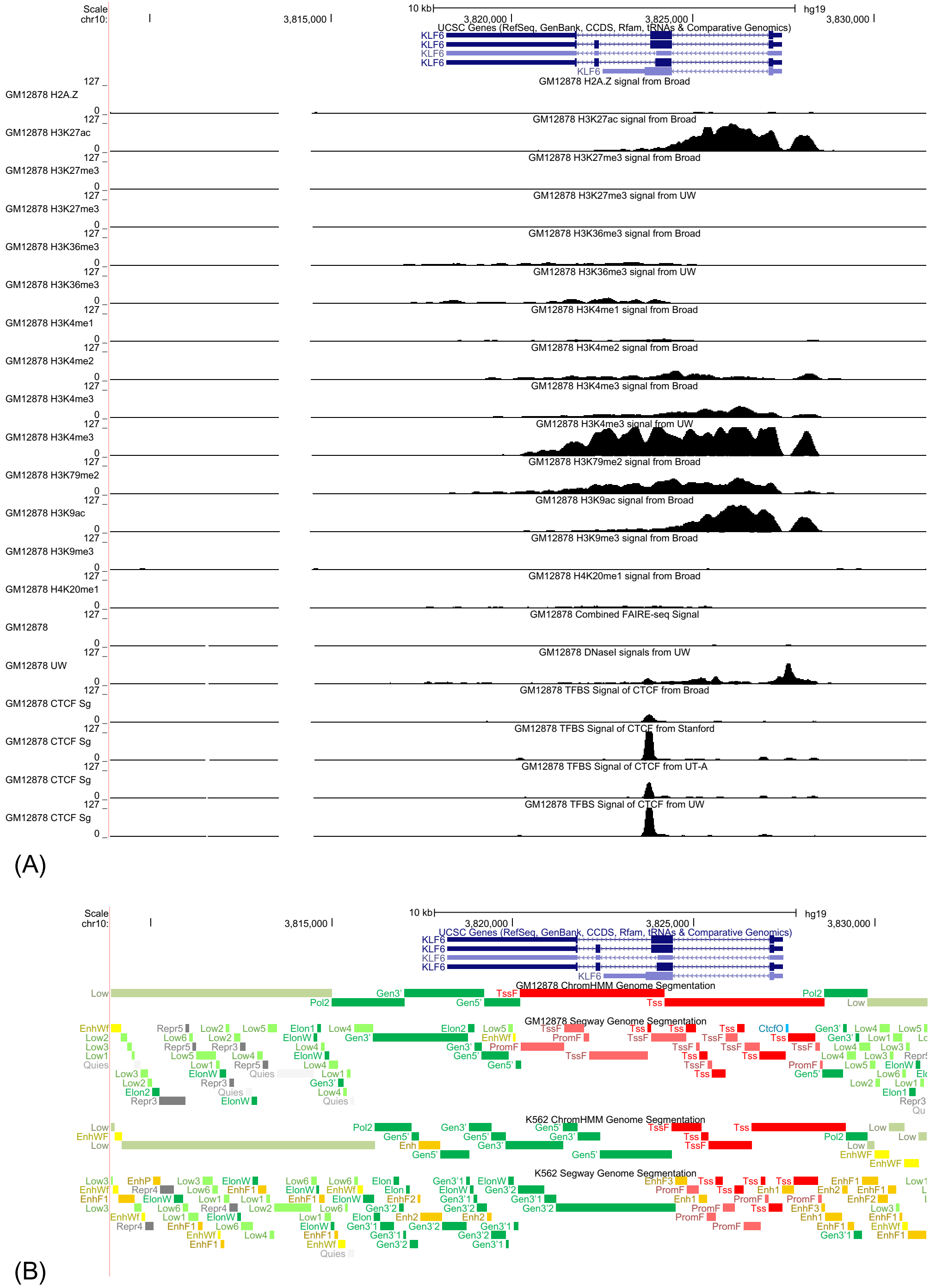


Fig. 1 Chromatin segmentation performed by unsupervised algorithms. (A) Signals from ChIP-seq, DNase-seq, and FAIRE for the GM12878 cell lines, from the ENCODE Consortium. (B) Segmentation tracks for GM12878 and K562 cell lines obtained with ChromHMM and Segway methods. All tracks have been obtained from ENCODE data repository at the UCSC Genome Browser. Color code for the segmentation tracks: bright red—active promoter, light red—promoter flanking, purple—inactive promoter, orange—candidate strong enhancer, yellow—candidate weak enhancer, blue—distal CTCF/candidate insulator, dark green—transcription associated, light green—low activity proximal to active states, gray—polycomb repressed, light gray—heterochromatin/repetitive/copy number variation.

DNse-seq, and similar experiments, and CENTIPEDE [35] uses Bayesian mixture models to cluster candidate TFBS based on open chromatin signal (from DNase-seq), presence of active histone marks and evolutionary conservation.[b]

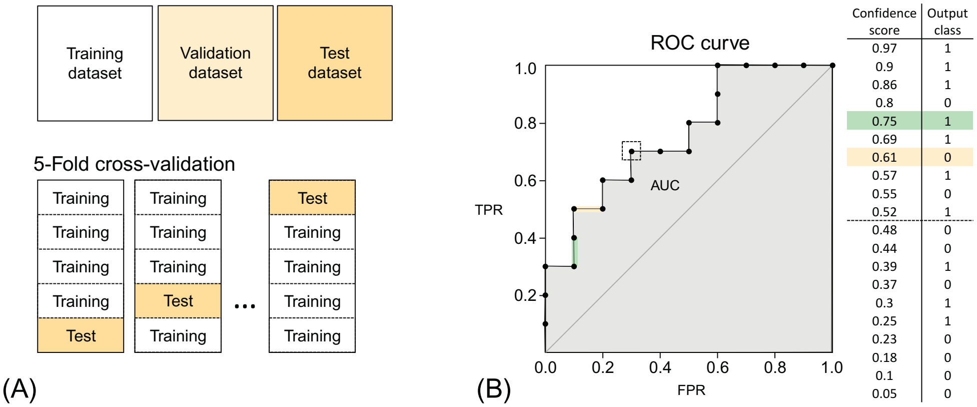
Supervised approaches

In contrast to the unsupervised machine learning approaches, the supervised classification algorithms require training datasets containing representative examples of all output classes. The final performance and usefulness of the classifier depend highly on the quality of the training dataset—the more representative, comprehensive, and unbiased the training data is, the better the classification results can be. The most common classification task in epigenomics is differentiation between active promoters or enhancers and other types of genomic elements. The optimal training dataset for this task should contain representatives of functional elements whose activity has been experimentally validated in vivo, ideally in relevant cell type or tissue (positive examples) and representatives of genomic intervals for which lack of enhancer activity was also shown (negative examples). The required size of the training dataset highly depends on the expected variability among both positive and negative cases, which partially results from the classification task formulation. Also the choice of the classification algorithm influences the required size of the training dataset and may limit the maximal number of tractable cases. Typically, researchers want to assemble datasets consisting of at least several hundred example regions, or preferably thousands of them for the trained model to be representative of the estimated tens or hundreds of thousands of actual regulatory regions. Two studies of predicting mesoderm specific enhancers serve as a good example of how data set size affects the accuracy. In the earlier study, it was very difficult to obtain a large number of examples and learning was done on relatively small data set of fewer than a hundred regions [36]; however, when more examples were available and the same methodology was applied to a set of over 8000 enhancers, the prediction accuracy was significantly higher [37]. These studies are described in more details in the "Prediction of enhancer regulatory state with Bayesian networks" section.

The assessment of the performance of the supervised classification model requires availability of a test dataset—a collection of labeled elements which were not used for model building (training). When high number of labeled cases are available, they can be divided into either two datasets: training dataset for model building and test dataset for model assessment, or additionally a third dataset, called the validation dataset, can be created and used for model parameters tuning before final assessment on the test

---

[b] For readers interested in seeing an actual code that goes into an unsupervised approach for chromatin state description, we refer them to the repository of the msCentipede method on github https://github.com/rajanil/msCentipede.



Confidence score	Output class
0.97	1
0.9	1
0.86	1
0.8	0
0.75	1
0.69	1
0.61	0
0.57	1
0.55	0
0.52	1
0.48	0
0.44	0
0.39	1
0.37	0
0.3	1
0.25	1
0.23	0
0.18	0
0.1	0
0.05	0

Fig. 2 Validation and assessment of supervised models. (A) Different strategies of splitting available datasets for use in training and testing of the model. (B) ROC curve. The curve can be plotted based on elements sorted by confidence score in decreasing order. Correctly classified element from class 1 results in one step up (example marked in green), while incorrectly classified element from class 0 in one step to the right (example marked in yellow). An example confidence score threshold = 0.5 (dotted line in the table and around one of the points) results in TPR = 0.7 and FPR = 0.3. Gray line represents ROC curve with AUC = 0.5. TPR—true-positive rate, FPR—false-positive rate, ROC—receiver operating characteristic, AUC—area under the curve.

set (Fig. 2A). Another approach, especially helpful if available datasets are small, is called cross-validation. In cross-validation, the available dataset is randomly divided into *n* parts (e.g., five in the case of fivefold cross-validation or 10 in 10-fold cross-validation) and model training and testing is repeated *n* times with one of the dataset parts used for testing and all other data used for training (Fig. 2A). The model performance is then usually expressed as the mean performance from the *n* testing procedures.

For a binary classifier, its performance can be measured in multiple ways. The most standard two are the true-positive rate (TPR, also called sensitivity), i.e., the ratio of correctly predicted positive examples to all positive examples and the true-negative rate (TNR, also called specificity), i.e., the ratio of the correctly predicted negative examples to all negative examples. These two measures also have their opposite counterparts (false-positive rate—FPR and false-negative rate—FNR) defined accordingly. It is important to note that while we can describe the classifier performance with different sets of two parameters (e.g., TPR and FPR), using only one of them (e.g., just reporting sensitivity) is not enough to judge the quality of predictions, as it is usually possible to have trivial models with very high sensitivity but with very low specificity or vice versa.

A slightly more complex measure often used to describe the ability of models to correctly distinguish between output classes is the area under the receiver operating characteristics (AUROC or AUC). It is only applicable to models that can provide a ranking of samples from the most "positive" to the most "negative" ones instead of a simple binary classification. Luckily, this is a broad class of models that includes probabilistic classification techniques (like Bayesian networks, where samples can be ranked by the posterior probability), voting techniques (including random forests, where the samples can be ranked by the positive vote fraction), and any weighted classifiers [including support vector machines (SVMs) and neural networks]. The measure is based on devising a curve based on the ranking (see Fig. 2B for details on how this is done for the ROC) and measuring the area under this curve. The higher the AUC value is, the better the model, with 1 being the maximal value, 0.5 meaning that the model is not capable of distinguishing between classes (equal to random guesses), and 0 being the minimal value, indicating that the model is always assigning the opposite class (the negative cases are labeled as positive and vice versa). Another, perhaps more intuitive, way to interpret the AUC measure is to observe that the value of AUC is equal to the probability of two samples, a negative and a positive one, randomly selected from the testing set to be ranked in the correct order (positive above the negative) by the classifier.

Methods for training data generation

Currently, the most reliable source of evidence of enhancer activity in vivo are transgenic reporter assays in which candidate enhancer DNA sequence is cloned upstream of a minimal promoter and reporter gene and the whole construct is stably integrated into the model organism genome (reviewed in Ref. [38]). This type of assay was performed on the scale of individual enhancers but the possibility to use results obtained with this method as training datasets for machine learning algorithms emerged when it was applied on a bigger scale. Pennacchio et al. have applied transgenic mouse reporter assays for functional testing of 167 putative human developmental enhancers which were computationally selected because of very high evolutionary sequence conservation [39]. Of these sequences, 45% have been reproducibly shown to function as tissue-specific enhancers of gene expression at embryonic day 11.5. The results of the assays, including stained whole-embryo images, have been collected in the VISTA Enhancer Browser [40].

An attempt to functionally characterize developmental enhancers on a genome-scale using transgenic *Drosophila melanogaster* lines was made by Kvon et al. [41]. Around 14 Mb (13.5%) of the noncoding regions of the fly genome have been divided into ~2 kb fragments representing enhancer candidates and integrated into identical position in the genome in constructs containing also minimal promoter and GAL4 reporter gene. Around 400 embryos representing all stages of development were analyzed for each transgenic line. Of the 7705 tested candidate enhancers 3557 have been found to be active in fly embryos and their activity patterns have been manually annotated, constituting around 5% of the total number of *Drosophila* developmental enhancers, which is estimated to be between 50,000 and 100,000.

To move enhancer assays to a truly genome-wide scale, an approach called STARR-seq (self-transcribing active regulatory region sequencing) has been developed [42]. In this method, candidate enhancer sequences are placed downstream of a minimal promoter such that active enhancers activate transcription of themselves. A genome-wide reporter library containing randomly sheared genomic DNA of the *D. melanogaster* reference strain and covering around 11.3 million candidate enhancer fragments has been cloned and used to transfect *Drosophila* S2 cells. Sequencing of polyadenylated RNA was performed and enrichment over input was quantified for each genomic position. As a result, 5499 candidates with enhancer activity have been identified. However, it should be noted that the increase in throughput gained by the application of STARR-seq compared with transgenic reporter assays may come together with the loss of reliability of results since STARR-seq is an episomal assay in which tested enhancers are not stably integrated into the genome. Even though subsets of enhancers identified by STARR-seq have been validated in assays involving stable integration into the genome, it is expected that such episomal assays may be associated with increased number false-positive and false-negative results, especially in mammalian cells [43].

Apart from the low- and high-throughput methods involving reporter assays that give direct evidence of enhancer activity in tested conditions, there is a number of experimental approaches providing indirect evidence of enhancer activity. Since active enhancers are expected to be associated with the presence of open chromatin regions distal to transcription start sites, high-throughput sequencing-based methods probing chromatin accessibility, such as ATAC-seq. [44], DNase-seq. [45], and FAIRE [46], are widely used to identify putative enhancer regions. Activity of enhancer elements has been also found to be associated with local transcription, giving rise to the so-called enhancer RNAs (eRNAs) [47], produced in opposing directions on each strand. This observation led to the use of methods measuring nascent RNA production like GRO-seq (as Global Run-On and sequencing) [48] and its modifications or capturing location of transcription initiation like CAGE [49] for the detection of enhancers activity [50, 51]. Each of these approaches not only provides indirect evidence of enhancer activity but also captures slightly different aspect of molecular processes associated with regulatory function of enhancer elements and thus it is expected that using results from each one of them as training data will be associated with introduction of different kinds of biases into the classification model.

Classical classification methods

Assuming that one has collected a training set with sufficient number of positive and negative examples (e.g., active enhancers and background genomic regions), each of which is coupled with a set of measured features (such as ChIP-seq measurements of histone modifications), one can attempt to apply any standard classification technique. We will review here a few examples of such approaches.

Prediction of enhancer regulatory state with Bayesian networks

Among popular and well-established supervised algorithms used for classification in the context of epigenomic studies, Bayesian networks [36], random forests [52, 53], and SVMs [54, 55] can be listed. Bayesian networks have been applied by Bonn et al. in an analysis of spatiotemporal activity of enhancers during *D. melanogaster* embryonic development [36]. The training set for this study comprised 65 enhancers from the CRM Activity Database 2 (CAD2), which were examined in vivo using transgenic reporter assays, located at least 1 kb from gene boundaries and having information about activity in mesoderm or outside mesoderm at 6–8 h of *Drosophila* embryo development. For this particular stage of development, quantitative measures of six chromatin marks (H3K4me3, H3K27ac, H3K79me3, H3K36me3, H3K4me1, and H3K27me3) and RNA polymerase II (Pol II) occupancy were collected. Each enhancer from the training set was represented as a vector of observed signals (input variables) and binary annotation for activity in or outside mesoderm and at selected time points (output label). The network with the maximal posterior probability given the data was found using the BNFinder package [56]. Conditional probability distributions of observing each activity class depending on the measured value of each input variable included in the network were also constructed. They represent conditional dependencies between inputs and outputs or network edges. Such a model could be used in the test or prediction phase to score unseen genomic loci by assigning them posterior probabilities of belonging to each activity class.

Two activity states were considered in this analysis of *Drosophila* enhancers: broader one, encompassing enhancers active in mesoderm at any stage of development, and more restricted one, in which only enhancers active in mesoderm at 6–8 h were considered. The model performance was assessed in a fourfold cross-validation scheme. The AUC for the “mesoderm” model was 0.76, while the “mesoderm at 6–8 h” model AUC was 0.82. The model has also identified input variables with predictive value for enhancer activity (H3K79me3, H3K27ac, and H3K27me3, the last one being negatively linked to the activity of the broader group of mesodermal enhancers). H3K4me1, H3K4me3, and H3K36me3 were found to have no predictive value for activity and Pol II was a causal dependency for mesodermal enhancers activity at 6–8 h of development. Overall, the model accurately represented both activity states; however, it should be noted that it was trained on a very small dataset. Finally, application of the trained network to genomic loci unseen during training resulted in de novo prediction of 112 enhancers expected to be active in mesoderm at 6–8 h of development, which significantly overlapped a collection of enhancers known to recruit mesodermal transcription factors at this stage of development. Nine of those predicted regions were tested for activity in vivo with transgenic reporter assays and eight of them were confirmed to be active in mesoderm at the expected stage of development. A simplified (with some variables removed for clarity) schematic representation of the Bayesian network approach to enhancer classification is presented in Fig. 3. In a later study by Podsiadlo et al. [37], it has been shown that the

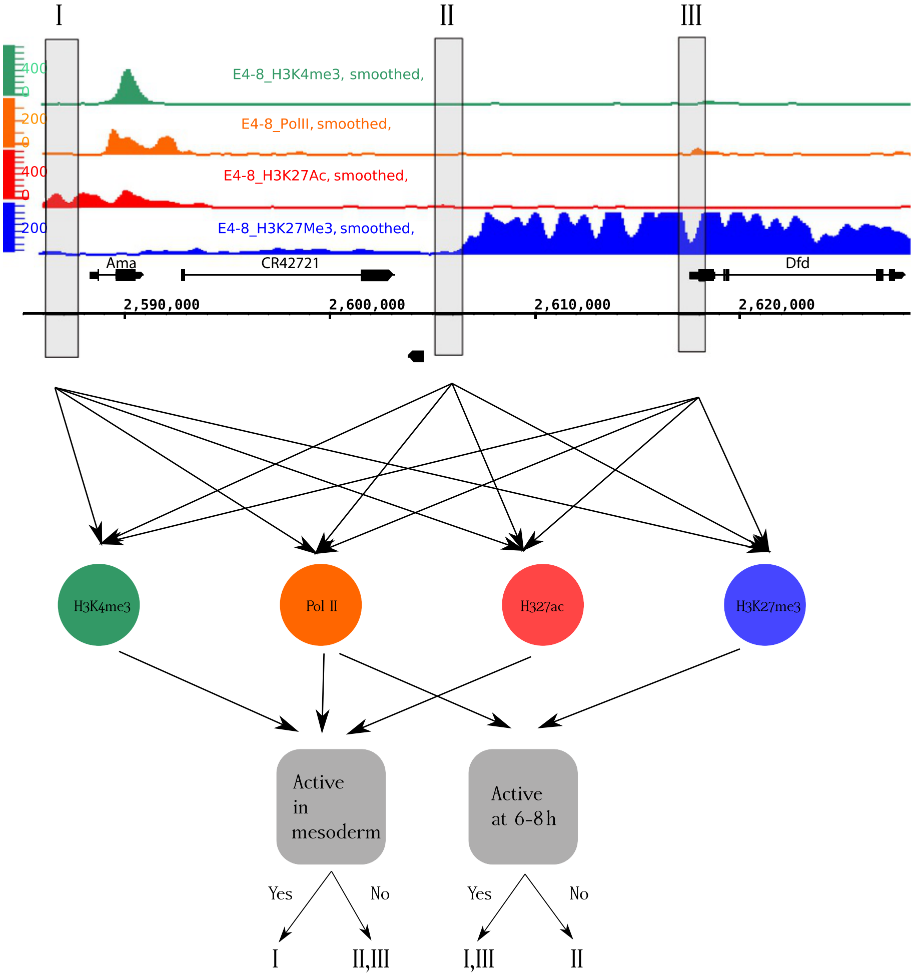

I
II
III
400
0
E4-8_H3K4me3, smoothed,
200
0
E4-8_PolII, smoothed,
400
0
E4-8_H3K27Ac, smoothed,
200
E4-8_H3K27Me3, smoothed,
Ama
CR42721
Dfd
2,590,000
2,600,000
2,610,000
2,620,000
H3K4me3
Pol II
H327ac
H3K27me3
Active in mesoderm
Active at 6-8 h
Yes
No
I
II,III
Yes
No
I,III
II

Fig. 3 A simplified schematic representation of the enhancer classification approach using Bayesian networks. Three different regions in the genome (I, II, III) display different profiles based on the same ChIP-seq data. All of these profiles are fed into the input variables (color coded after the ChIP-seq data) of the Bayesian network and converted to posterior probabilities of activity in mesoderm (left output variable) and at 6–8 h of development (right output variable). Each of the three exemplary regions are classified differently by the network.

same approach of predicting mesodermal enhancers in fruit flies can be significantly improved to AUC of over 90% by increasing the size of the training set to over 8008 putative enhancers defined by the binding patterns of mesodermal TFs.[c]

Multiple kernel learning approach for the identification of tissue specific developmental enhancers

Active enhancers are expected to be characterized not only by specific distribution of histone marks but also by sequence features and evolutionary conservation. To be able to integrate multiple data types into a single discrimination function, authors of the EnhancerFinder method [54] have used multiple kernel learning (MKL), which is an extension of SVM. Three data types have been integrated into EnhancerFinder: (i) "functional genomics data" which encompass the presence of several histone modifications, protein-DNA associations for TFs and p300 protein, and several measurements of open chromatin (represented in a binary format based on peak calling results available within the ENCODE data repository at the UCSC Genome Browser); (ii) DNA sequence motif patterns (number of occurrences of all possible 4-mers in the sequence); and (iii) evolutionary conservation based on mammalian phastCons elements [57] (maximal score within a genomic region). Each of them is represented in the MKL model by one kernel function with weighted contribution.

Prediction of tissue-specific developmental enhancers by EnhancerFinder is divided into two steps: first developmental enhancers are distinguished from genomic background, then tissue specificity prediction is performed for identified putative enhancers. At both stages, training data originate from the VISTA Enhancer Browser [40]—a catalogue of mammalian enhancers with in vivo validated activity. Training data for the first step include 711 VISTA enhancers active in any tissue as positive examples and 711 random genomic regions with matched length and chromosome distributions as negative examples. In the second step, tissue-specific subsets of 1447 VISTA regions are used (for each tissue regions active in it are positive examples while regions active only in other tissues or not active at all are negative examples). Algorithms performance was evaluated using 10-fold cross-validation to compute the AUROC value. EnhancerFinder discriminated between active VISTA enhancers and genomic background very efficiently (AUROC 0.96) and exhibited significant advantage when compared with SVM classifiers using only one of the three data types. The performance of the tissue specificity identification step was lower and differed highly between the analyzed tissues, with the highest value being 0.85 for heart, followed by 0.74 for limb, 0.72 for forebrain and midbrain, 0.69 for hindbrain, and 0.62 for neural tube.

---

[c] Readers interested in learning how to apply Bayesian networks for classification can see some code examples in the BNfinder tutorial at the github repository https://github.com/sysbio-vo/bnfinder/.

The two-step procedure of tissue-specific enhancers identification was shown to allow for more precise prediction of tissue specificity compared to approaches aiming at performing prediction of activity of an enhancer and its tissue specificity in one step. Comparison of predicted tissue distributions to those of validated enhancers from the VISTA database revealed that one-step approaches tend to miss tissue specificity and predict very similar sets of enhancers regardless of the target tissue.

EnhancerFinder was also applied for de novo prediction of developmental enhancers in the human genome. Due to the fact that the vast majority of positive training examples exhibited high levels of evolutionary conservation, the information about conservation was not included in the first step of the prediction, to increase the method ability to detect less conserved novel enhancers. Overall, around 84,000 developmental enhancers have been predicted at 5% FPR threshold, of which 7400 were identified as possible limb enhancers, 19,051 as heart enhancers, and 11,693 as brain enhancers. These genome-wide predictions were characterized by enrichment near genes with annotated roles in their predicted tissues and presence of high number of lead SNPs from genome-wide association studies. Selected predictions were validated in vivo in transgenic mice model and zebrafish.

Prediction of active enhancers based on DNA methylation marks and histone modifications with random forest classifier

Most enhancer prediction methods depend predominantly on signal from histone modifications. These signals have, however, relatively low resolution, resulting from the resolution of ChIP-seq methodology. Since depletion in DNA methylation in the mCG context is expected to be a signature of active enhancers [58, 59], authors of the REPTILE method used data from whole-genome bisulfite sequencing, which informs about DNA methylation in 1 bp resolution, to improve the resolution of enhancer prediction [52]. They emphasized that genomic regions typically labeled as enhancers are too wide compared to the actual TF-binding sites and as a result contain a big fraction of sequence without regulatory potential. To point possible functional regulatory subregions within such wide intervals, they identified differentially methylated regions (DMRs)—genomic intervals characterized by difference in mCG between target sample and reference samples from different tissue/cell types. Then REPTILE uses two random forest (RF) classifiers to distinguish enhancers from genomic background. The first RF models epigenomic signatures of positive and negative enhancer examples. The second RF captures epigenomic signatures of DRMs located in active and inactive enhancers. Each DRM and query region is represented by a feature vector, which stores a pair of values for each epigenetic mark: intensity and intensity deviation calculated with respect to reference samples. In the prediction step, the confidence score assigned to the candidate enhancer region is the maximum of the score assigned to the complete region by the first RF and scores assigned by the second RF to all DRMs located within this region.

REPTILE was trained to predict human and mouse enhancers using epigenomic datasets from H1 human embryonic stem cells and mouse embryonic stem cells, respectively. EP300 binding data from human and mouse ESCs were used as putative active enhancers (positive examples) while promoters and random genomic regions were used as negative examples. For human, four H1-derived cell type data were included as reference. For mouse, data from eight tissues from embryonic day 11.5 (E11.5) were used as reference. The epigenomic signatures of analyzed elements included six histone modifications: H3K4me1, H3K4me2, H3K4me3, H3K27me3, H3K27ac, and H3K9ac and mCG information. REPTILE outperformed four other methods trained on the same datasets (PEDLA [60], RFECS [61], DELTA [62], CSIANN [63]) when the predictions were validated against experimentally verified enhancers from the VISTA database (for mouse) or publicly available datasets describing DNase hypersensitivity sites (for human). Additionally, predictions made by REPTILE exhibited highest enrichment in TF motifs relevant for cell/tissue type for which the prediction was performed, however, the authors noted significant overlaps between predictions in different tissues.

Some of these issues were tackled by the study of Herman-Izycka et al. [53], which also used random forest classifiers to predict active enhancers in different tissues based on the same VISTA dataset. In this case, the problem of high overlap between predictions of classifiers for different tissues was addressed by using a two-level classifier that was based on two random forests, one trained on enhancers from the targeted tissue, and the other trained on a set of non-tissue–specific promoters. As the study showed, the addition of the second classifier, which was specifically suited to filter out the nonspecific signal associated with general transcription activation, was able to alleviate the issue of cross-predicting enhancers from another, non-related tissue.

New approaches—Deep learning

In recent years, we have witnessed a renaissance of artificial neural networks as a tool that has given a considerable advantage over the classical machine learning approaches in many applications that were considered standard, with the most notable example of image recognition in general and face recognition in particular. The field of epigenomics has also seen an advent of neural network (NN) approaches that were tried in the context of the same datasets as we have discussed in the previous section.

While these new architectures of neural networks can be all attached to the generic term of "deep learning" based on the fact that they usually consist of large number (dozens) of consecutively connected layers, there are, however, additional distinctions of subtypes of such networks. In particular, many successful approaches employ convolutional neural networks (cNN) that are characterized by the usage of many, not fully connected, convolutional layers that are closely analogous to the image processing filters. They are very successful in feature extraction from images; however, as it was shown in many applications to text or number processing, this idea has many applications beyond image analysis. Another frequently employed architecture are the so-called recurrent neural networks (rNN) that are characterized by sequential use of similar layers connected in a way that mimics long- and short-term memory in biological neural systems. This usually involves some sort of cyclic connections between layers.[d]

It should be noted that all deep learning approaches require usually more data points to properly fit a model than the more classical supervised learning techniques such as random forests. While a typical dataset for classical machine learning would include hundreds of examples (with some cases like Bonn et al. [36] where a successful model was built on dozens of examples), it is not uncommon in deep learning approaches to train on tens or hundreds of thousands of examples. In some tasks, like image recognition or computer games, there is no difficulty in supplying such large numbers of training examples. However, in epigenomics, it is not typically feasible to provide so many well validated biological examples. This is why, in the context of predicting epigenomic state some compromises have to be made. There are practically two ways that are employed toward this end—either the neural network is used on a relatively small dataset as in the case of BiRen [65] method that is trained on the Vista enhancers, which is relatively small for neural networks, or in other approaches the training set is extended to contain thousands of examples, but this is at the cost of utilizing some of the less direct assays such as DNase-seq, as it is in the case of Basset method [66].

Historically, the first study to use deep learning methodology in the context of epigenomics was the CSI-ANN [63]. In this approach, the authors have applied a hybrid model that used a Fisher discriminant analysis method coupled with a time delay NN model to predict a set of putative enhancers (classified based on p300 binding) using subsets of histone modifications as features. As it usually is with pioneering studies, the results were far from perfect. The training set was very small (only 213 enhancers) and the performance is also not ideal (sensitivity and positive predictive value near 65%).

After the CSI-ANN method was published, there were several other approaches that used a similar premise of attempting to predict binding patterns of different proteins based on large-scale data with a neural network as a classification device. The DeepBind method [67] extended this approach to almost a thousand of different DNA binding proteins and used cNNs to predict binding with almost 100% accuracy. Later, Zeng and Gifford [68] provided an overview of different cNN architectures that can be used in this task and showed that similar performance can be achieved with much simpler networks consisting of 1, 2, or 3 convolutional layers. These results were further improved by the authors of the KEGRU [69] method that combined convolutional with recurrent networks and gave an even higher prediction accuracy on the ENCODE ChIP-seq binding data than DeepBind and other previous approaches.

---

[d] For more details regarding the neural networks construction we would like to refer the readers to the Deep Learning book by Goodfellow, Bengio, and Courvillle [64].

More recently, the rNNs were used to accurately predict the DNase hypersensitivity patterns based on the DNA sequence [70]. This work extends on an earlier approach called Basset [66], which solved the same problem of predicting DNase data based on DNA sequence but with cNNs, similar to the Deopen method developed slightly later [71].[e]

Another set of interesting approaches are using neural networks to predict impact of single-point mutations on either histone modifications (DeepSEA [72]) or DNA methylation (CpGenie [73]). While the DeepSEA framework uses a deep NN architecture, the CpGenie is based on the convolutional NN paradigm. Both are trained on single nucleotide polymorphic (SNP) sites to identify which of the SNPs are likely to have an impact on the epigenomic state of the locus. While in both cases the performance so far is not perfect, they are showing clearly that the neural-network based approaches have the capability of addressing questions (like role of mutations) that were unreachable by classical ML methods.

Another creative use of neural networks in epigenomics are approaches like Coda [74], which are aimed at increasing the quality of genome wide datasets. In this particular case, the method is trained on poor-quality datasets together with higher quality data to be able to detect peaks and other signals in lower quality unseen data. This trend of attempts to improve experimental data quality is also used in the hic-net [75] software for processing Hi-C data.

Conclusion

The last decade has brought us wealth of epigenomic data both in terms of the high-quality validated functional datasets and high-throughput genome wide data, which is usually of much lower fidelity.

Machine learning techniques have proven to be very useful in this area. First, we can clearly use them for processing raw data giving us automatic annotations of the genome like the ChromHMM and similar unsupervised methods. Second, the classical supervised machine learning techniques, including SVMs and Random Forests, can be reliably used to generalize the information we have in middle to large-scale functional data (like reporter assays) on the basis of the genome wide measurements (such as the histone modifications). Third, the advent of different kinds of deep learning methods is already making an impact on the ways in which we analyze epigenomic data; however, given that most of the discussed methods appeared in the last 5 years, we think that the time will tell which of these approaches will prove to be the most useful in the long run.

---

[e] Readers interested in reviewing an example code for a deep learning-based approach can see the Deopen method github repository and its documentation https://github.com/kimmo1019/Deopen.

Acknowledgments

This study was supported by the POWROTY/REINTEGRATION program of the Foundation for Polish Science cofinanced by the European Union under the European Regional Development Fund (MM) and the Polish National Science Center grant [DEC 2015/16/W/NZ2/00314] (BW).

References

[1] Wilczynski B, Furlong EEM. Challenges for modeling global gene regulatory networks during development: insights from Drosophila. Dev Biol 2010;340:161–9. https://doi.org/10.1016/j.ydbio.2009.10.032.

[2] Bird A. DNA methylation patterns and epigenetic memory. Genes Dev 2002;16:6–21. https://doi.org/10.1101/gad.947102.

[3] Barski A, Cuddapah S, Cui K, Roh T-Y, Schones DE, Wang Z, et al. High-resolution profiling of histone methylations in the human genome. Cell 2007;129:823–37. https://doi.org/10.1016/j.cell.2007.05.009.

[4] Woodcock CL, Ghosh RP. Chromatin higher-order structure and dynamics. Cold Spring Harb Perspect Biol 2010;2. https://doi.org/10.1101/cshperspect.a000596. a000596.

[5] Csankovszki G, Nagy A, Jaenisch R. Synergism of Xist Rna, DNA methylation, and histone Hypoacetylation in maintaining X chromosome inactivation. J Cell Biol 2001;153:773–84. https://doi.org/10.1083/jcb.153.4.773.

[6] Cameron EE, Bachman KE, Myöhänen S, Herman JG, Baylin SB. Synergy of demethylation and histone deacetylase inhibition in the re-expression of genes silenced in cancer. Nat Genet 1999;21:103–7. https://doi.org/10.1038/5047.

[7] Chiurazzi P, Grazia Pomponi M, Pietrobono R, Bakker CE, Neri G, Oostra BA. Synergistic effect of histone Hyperacetylation and DNA Demethylation in the reactivation of the FMR1 gene. Hum Mol Genet 1999;8:2317–23. https://doi.org/10.1093/hmg/8.12.2317.

[8] Li L, Lyu X, Hou C, Takenaka N, Nguyen HQ, Ong C-T, et al. Widespread rearrangement of 3D chromatin organization underlies Polycomb-mediated stress-induced silencing. Mol Cell 2015;58:216–31. https://doi.org/10.1016/j.molcel.2015.02.023.

[9] Didonna A, Opal P. The promise and perils of HDAC inhibitors in neurodegeneration. Ann Clin Transl Neurol 2015;2:79–101. https://doi.org/10.1002/acn3.147.

[10] McCulloch WS, Pitts W. A logical calculus of the ideas immanent in nervous activity. Bull Math Biophys 1943;5:115.

[11] Rosenblatt F. The perceptron, a perceiving and recognizing automaton project para. Cornell Aeronautical Laboratory; 1957.

[12] Turing AM. I.—Computing machinery and intelligence. Mind 1950;LIX:433–60. https://doi.org/10.1093/mind/LIX.236.433.

[13] Holtzman S. Intelligent decision systems. New York: Addison-Wesley; 1989.

[14] Lemon SC, Roy J, Clark MA, Friedmann PD, Rakowski W. Classification and regression tree analysis in public health: methodological review and comparison with logistic regression. Ann Behav Med 200326:172–81.

[15] Dietterich T. Overfitting and undercomputing in machine learning. ACM Comput Surv 1995;27:326–7.

[16] Hastie T, Tibshirani R, Friedman J. The elements of statistical learning; data mining, inference and prediction. 2nd ed. New York: Springer New York; 2009.

[17] Breiman L. Random forests. Mach Learn 2001;45:5–32.

[18] Efron B, Tibshirani R. Improvements on cross-validation: the 632+ bootstrap method. J Am Stat Assoc 1997;92:548–60. https://doi.org/10.1080/01621459.1997.10474007.

[19] Rouzier R. Breast Cancer molecular subtypes respond differently to preoperative chemotherapy. Clin Cancer Res 2005;11:5678–85. https://doi.org/10.1158/1078-0432.CCR-04-2421.

[20] Sotiriou C, Neo S-Y, McShane LM, Korn EL, Long PM, Jazaeri A, et al. Breast cancer classification and prognosis based on gene expression profiles from a population-based study. Proc Natl Acad Sci 2003;100:10393–8. https://doi.org/10.1073/pnas.1732912100.

[21] Beer DG, Kardia SLR, Huang C-C, Giordano TJ, Levin AM, Misek DE, et al. Gene-expression profiles predict survival of patients with lung adenocarcinoma. Nat Med 2002;8:816–24. https://doi.org/10.1038/nm733.

[22] Kim E, Oh W, Pieczkiewicz DS, Castro MR, Caraballo PJ, Simon GJ. Divisive hierarchical clustering towards identifying clinically significant pre-diabetes subpopulations. In: AMIA 2014 Annu Symp; 2014.

[23] Wang C, Gong L, Yu Q, Li X, Xie Y, Zhou X. DLAU: a scalable deep learning accelerator unit on FPGA. IEEE Trans Comput Des Integr Circuits Syst 2017;36:513–7.

[24] Abadi M, Barham P, Chen J, Chen Z, Davis A, Dean J, et al. Tensorflow: A system for large-scale machine learning. In: 12th {USENIX} symposium on operating systems design and implementation ({OSDI} 16); 2016. p. 265–83.

[25] LeCun Y, Bengio Y, Hinton G. Deep learning. Nature 2015;521:436.

[26] Krizhevsky A, Sutskever I, Hinton GE. Imagenet classification with deep convolutional neural networks. In: Advances in neural information processing systems; 2012. p. 1097–105.

[27] Dojer N, Bednarz P, Podsiadło A, Wilczyński B. BNFinder2: faster Bayesian network learning and Bayesian classification. Bioinformatics 2013;29:2068–70. https://doi.org/10.1093/bioinformatics/btt323.

[28] Cortes C, Vapnik V. Support-vector networks. Mach Learn 1995;20:273.

[29] Ernst J, Kellis M. ChromHMM: automating chromatin-state discovery and characterization. Nat Methods 2012;9:215–6. https://doi.org/10.1038/nmeth.1906.

[30] Ernst J, Kellis M. Discovery and characterization of chromatin states for systematic annotation of the human genome. In: Lecture Notes in Computer Science (including Subser Lect Notes Artif Intell Lect Notes Bioinformatics), vol. 6577; 2011. https://doi.org/10.1007/978-3-642-20036-6_6. LNBI:53.

[31] Hoffman MM, Ernst J, Wilder SP, Kundaje A, Harris RS, Libbrecht M, et al. Integrative annotation of chromatin elements from ENCODE data. Nucleic Acids Res 2013;41:827–41.

[32] Filion GJ, van Bemmel JG, Braunschweig U, Talhout W, Kind J, Ward LD, et al. Systematic protein location mapping reveals five principal chromatin types in Drosophila cells. Cell 2010;143:212–24.

[33] Hoffman MM, Buske OJ, Wang J, Weng Z, Bilmes JA, Noble WS. Unsupervised pattern discovery in human chromatin structure through genomic segmentation. Nat Methods 2012;9:473–6. https://doi.org/10.1038/nmeth.1937.

[34] Rashid NU, Giresi PG, Ibrahim JG, Sun W, Lieb JD. ZINBA integrates local covariates with DNA-seq data to identify broad and narrow regions of enrichment, even within amplified genomic regions. Genome Biol 2011;12:R67. https://doi.org/10.1186/gb-2011-12-7-r67.

[35] Pique-Regi R, Degner JF, Pai AA, Gaffney DJ, Gilad Y, Pritchard JK. Accurate inference of transcription factor binding from DNA sequence and chromatin accessibility data. Genome Res 2011;21:447–55. https://doi.org/10.1101/gr.112623.110.

[36] Bonn S, Zinzen RP, Girardot C, Gustafson EH, Perez-Gonzalez A, Delhomme N, et al. Tissue-specific analysis of chromatin state identifies temporal signatures of enhancer activity during embryonic development. Nat Genet 2012;44:148–56. https://doi.org/10.1038/ng.1064.

[37] Podsiadło A, Wrzesień M, Paja W, Rudnicki W, Wilczyński B. Active enhancer positions can be accurately predicted from chromatin marks and collective sequence motif data. BMC Syst Biol 2013; 7(Suppl 6):1–7.

[38] Kvon EZ. Using transgenic reporter assays to functionally characterize enhancers in animals. Genomics 2015;106:185–92. https://doi.org/10.1016/j.ygeno.2015.06.007.

[39] Pennacchio LA, Ahituv N, Moses AM, Prabhakar S, Nobrega MA, Shoukry M, et al. In vivo enhancer analysis of human conserved non-coding sequences. Nature 2006;444:499–502. https://doi.org/10.1038/nature05295.

[40] Visel A, Minovitsky S, Dubchak I, Pennacchio LA. VISTA enhancer browser—a database of tissue-specific human enhancers. Nucleic Acids Res 2007;35(Database):D88–92. https://doi.org/10.1093/nar/gkl822.

[41] Kvon EZ, Kazmar T, Stampfel G, Yáñez-Cuna JO, Pagani M, Schernhuber K, et al. Genome-scale functional characterization of Drosophila developmental enhancers in vivo. Nature 2014;512:91–5.

[42] Arnold CD, Gerlach D, Stelzer C, Boryn LM, Rath M, Stark A. Genome-wide quantitative enhancer activity maps identified by STARR-seq. Science 2013;339:1074–7. https://doi.org/10.1126/science.1232542.

[43] Muerdter F, Boryń ŁM, Woodfin AR, Neumayr C, Rath M, Zabidi MA, et al. Resolving systematic errors in widely used enhancer activity assays in human cells. Nat Methods 2018;15:141–9. https://doi.org/10.1038/nmeth.4534.

[44] Buenrostro JD, Giresi PG, Zaba LC, Chang HY, Greenleaf WJ. Transposition of native chromatin for fast and sensitive epigenomic profiling of open chromatin, DNA-binding proteins and nucleosome position. Nat Methods 2013;10:1213–8. https://doi.org/10.1038/nmeth.2688.

[45] Song L, Crawford GE. DNase-seq: a high-resolution technique for mapping active gene regulatory elements across the genome from mammalian cells. Cold Spring Harb Protoc 2010. https://doi.org/10.1101/pdb.prot5384.

[46] Giresi PG, Kim J, McDaniell RM, Iyer VR, Lieb JD. FAIRE (formaldehyde-assisted isolation of regulatory elements) isolates active regulatory elements from human chromatin. Genome Res 2007;17:877–85. https://doi.org/10.1101/gr.5533506.

[47] Kim T-K, Hemberg M, Gray JM, Costa AM, Bear DM, Wu J, et al. Widespread transcription at neuronal activity-regulated enhancers. Nature 2010;465:182–7. https://doi.org/10.1038/nature09033.

[48] Core LJ, Waterfall JJ, Lis JT. Nascent RNA sequencing reveals widespread pausing and divergent initiation at human promoters. Science 2008;322:1845–8. https://doi.org/10.1126/science.1162228.

[49] Kodzius R, Kojima M, Nishiyori H, Nakamura M, Fukuda S, Tagami M, et al. CAGE: cap analysis of gene expression. Nat Methods 2006;3:211–22. https://doi.org/10.1038/nmeth0306-211.

[50] Core LJ, Martins AL, Danko CG, Waters CT, Siepel A, Lis JT. Analysis of nascent RNA identifies a unified architecture of initiation regions at mammalian promoters and enhancers. Nat Genet 2014;46:1311–20. https://doi.org/10.1038/ng.3142.

[51] Andersson R, Gebhard C, Miguel-Escalada I, Hoof I, Bornholdt J, Boyd M, et al. An atlas of active enhancers across human cell types and tissues. Nature 2014;507:455–61. https://doi.org/10.1038/nature12787.

[52] He Y, Gorkin DU, Dickel DE, Nery JR, Castanon RG, Lee AY, et al. Improved regulatory element prediction based on tissue-specific local epigenomic signatures. Proc Natl Acad Sci 2017;114:E1633–40. https://doi.org/10.1073/pnas.1618353114.

[53] Herman-Izycka J, Wlasnowolski M, Wilczynski B. Taking promoters out of enhancers in sequence based predictions of tissue-specific mammalian enhancers. BMC Med Genomics 2017;10(Suppl. 1):17–26. https://doi.org/10.1186/s12920-017-0264-3.

[54] Erwin GD, Oksenberg N, Truty RM, Kostka D, Murphy KK, Ahituv N, et al. Integrating diverse datasets improves developmental enhancer prediction. PLoS Comput Biol 2014;10:e1003677. https://doi.org/10.1371/journal.pcbi.1003677.

[55] Fletez-Brant C, Lee D, McCallion AS, Beer MA. Kmer-SVM: a web server for identifying predictive regulatory sequence features in genomic data sets. Nucleic Acids Res 2013;41:W544–56. https://doi.org/10.1093/nar/gkt519.

[56] Wilczynski B, Dojer N. BNFinder: exact and efficient method for learning Bayesian networks. Bioinformatics 2009;25:286–7. https://doi.org/10.1093/bioinformatics/btn505.

[57] Siepel A, Bejerano G, Pedersen JS, Hinrichs AS, Hou M, Rosenbloom K, et al. Evolutionarily conserved elements in vertebrate, insect, worm, and yeast genomes. Genome Res 2005;15:1034–50.

[58] Hwang W, Oliver VF, Merbs SL, Zhu H, Qian J. Prediction of promoters and enhancers using multiple DNA methylation-associated features. BMC Genomics 2015;16(Suppl. 7):S11. https://doi.org/10.1186/1471-2164-16-S7-S11.

[59] Stadler MB, Murr R, Burger L, Ivanek R, Lienert F, Schöler A, et al. DNA-binding factors shape the mouse methylome at distal regulatory regions. Nature 2011;480:490–5. https://doi.org/10.1038/nature10716.

[60] Liu F, Li H, Ren C, Bo X, Shu W. PEDLA: predicting enhancers with a deep learning-based algorithmic framework. Sci Rep 2016;6:1–14. https://doi.org/10.1038/srep28517.

[61] Rajagopal N, Xie W, Li Y, Wagner U, Wang W, Stamatoyannopoulos J, et al. RFECS: a random-forest based algorithm for enhancer identification from chromatin state. PLoS Comput Biol 2013;9e1002968.

[62] Lu Y, Qu W, Shan G, Zhang C. DELTA: a distal enhancer locating tool based on AdaBoost algorithm and shape features of chromatin modifications. PLoS ONE 2015;10:e0130622. https://doi.org/10.1371/journal.pone.0130622.

[63] Firpi HA, Ucar D, Tan K. Discover regulatory DNA elements using chromatin signatures and artificial neural network. Bioinformatics 2010;26:1579–86.

[64] Goodfellow I, Bengio Y, Courville A. Deep learning. Cambridge, MA: MIT Press; 2016.

[65] Yang B, Liu F, Ren C, Ouyang Z, Xie Z, Bo X, et al. BiRen: predicting enhancers with a deep-learning-based model using the DNA sequence alone. Bioinformatics 2017;33:1930–6.

[66] Kelley DR, Snoek J, Rinn JL. Basset: learning the regulatory code of the accessible genome with deep convolutional neural networks. Genome Res 2016;26:990–9. https://doi.org/10.1101/gr.200535.115.

[67] Alipanahi B, Delong A, Weirauch MT, Frey BJ. Predicting the sequence specificities of DNA- and RNA-binding proteins by deep learning. Nat Biotechnol 2015;33:831–8. https://doi.org/10.1038/nbt.3300.

[68] Zeng H, Edwards MD, Liu G, Gifford DK. Convolutional neural network architectures for predicting DNA-protein binding. Bioinformatics 2016;32:i121–7.

[69] Shen Z, Bao W, Huang D-S. Recurrent neural network for predicting transcription factor binding sites. Sci Rep 2018;8:15270. https://doi.org/10.1038/s41598-018-33321-1.

[70] Min X, Zeng W, Chen N, Chen T, Jiang R. Chromatin accessibility prediction via convolutional long short-term memory networks with k-mer embedding. Bioinformatics 2017;33:i92–101. https://doi.org/10.1093/bioinformatics/btx234.

[71] Liu Q, Xia F, Yin Q, Jiang R. Chromatin accessibility prediction via a hybrid deep convolutional neural network. Bioinformatics 2018;34:732–8. https://doi.org/10.1093/bioinformatics/btx679.

[72] Zhou J, Troyanskaya OG. Predicting effects of noncoding variants with deep learning-based sequence model. Nat Methods 2015;12:931–4.

[73] Zeng H, Gifford DK. Predicting the impact of non-coding variants on DNA methylation. Nucleic Acids Res 2017;45. https://doi.org/10.1093/nar/gkx177. e99-e99.

[74] Koh PW, Pierson E, Kundaje A. Denoising genome-wide histone ChIP-seq with convolutional neural networks. Bioinformatics 2017;33:i225–33. https://doi.org/10.1093/bioinformatics/btx243.

[75] Öztürk Ş, Akdemir B. HIC-net: a deep convolutional neural network model for classification of histopathological breast images. Comput Electr Eng 2019;76:299–310. https://doi.org/10.1016/j.compeleceng.2019.04.012.

This page intentionally left blank

CHAPTER 10

Systems immunology meets epigenetics

Wenhui Li[a,b,*], **Ziyi Chen**[a,b,*], **Aiping Wu**[a,b], **F. Xiao-Feng Qin**[a,b], **Lianjun Zhang**[a,b]
[a]Suzhou Institute of Systems Medicine, Suzhou, Jiangsu, China
[b]Center for Systems Medicine, Institute of Basic Medical Sciences, Chinese Academy of Medical Sciences and Peking Union Medical College, Beijing, China

Contents

Epigenetic modifications within the immune system 239
  DNA methylation 239
  RNA modification 242
  Histone modifications 243
Systems approach for deconvoluting immune cell composition 243
  Deconvolution frameworks 244
  Reference-based models 244
  Reference-free models 247
  Perspectives 248
References 250

Epigenetic modifications within the immune system

Epigenetic modifications are defined as stable and heritable alterations in gene expression and cellular function without altering the sequence of nucleotides. In general, epigenetic modifications can be divided into several categories including DNA methylation (DNAm), histone modification, DNA accessibility, chromatin structure alteration, and noncoding RNAs (ncRNAs). Epigenetic modifications in immune cells involve important changes in gene activity to dictate immune cell fate decisions and fulfill their biological functions. In previous epigenome-wide association studies (EWAS), DNA methylation was shown to be the most relevant epigenetic modification. By adding a methyl (–CH3) group at cytosines of CG dinucleotides, it can inhibit the transcription of genes and thus affect the cellular functions. Most cell types display relatively stable DNAm patterns, with 70%–80% of all CpGs being methylated.

DNA methylation

DNA methylation plays critical roles in immune regulation by modulating gene expression patterns, which is mainly controlled by DNA methyltransferases and methyl-CpG binding factors [1, 2]. CpG island, located mainly in gene promoter regions, is

---

[*] These authors contribute equally to this manuscript.

Epigenetics of the Immune System
https://doi.org/10.1016/B978-0-12-817964-2.00010-1
© 2020 Elsevier Inc.
All rights reserved.

predominantly methylated and affects chromatin accessibility and restricts transcription factors docking. Methylated CpG could attract methyl-CpG binding factors and recruit repressor complexes to alter the chromatin structure and suppress target gene expression.

Myeloid lineages may rapidly respond to environmental stimuli via a variety of pattern recognition receptors. Recently, it was well shown that the differentiation and functions of myeloid lineages can be regulated by epigenetic mechanisms. Monocyte, the progenitor of macrophage and dendritic cells (DC), is the main producer of interleukin-1β cytokine. IL1β promoter is highly accessible in differentiated monocytic cells compared to primary monocytes. The poised conformation of IL1β promoter is opened during monopoiesis, and demethylation leads to the induction of IL1β expression [3]. In human sepsis, tolerized monocytes harbor distinct DNA methylation patterns in its CpG island compared with healthy controls. For instance, the expression of IL10 and IL6 is significantly increased in sepsis monocytes, which was characterized by hypomethylation of their individual CpG sites. Additionally, the demethylation of CpG islands may also affect specific chromatin features to favor the transcription factor binding motif to enhance gene expression [4]. Similarly, during the differentiation of human monocytes into dendritic cells, epigenetic changes take place to regulate CD14 and CD209 expression. Marked demethylation occurred at the CD209 promoter CpG islands (CpG2 and CpG3), which is associated with enhanced transcription upon differentiation compared to hypermethylation of CD14 promoter [5].

Moreover, DNA methylation also regulates macrophage differentiation and functions [6]. Macrophages can be generally divided into M1 (classically activated) and M2 (alternatively activated) based on differential phenotypic or functional characteristics. A recent study shows that DNA methylation modulates macrophage differentiation/polarization at the feto-maternal interface. They found that M1-macrophage signature encoding genes were hypermethylated, whereas the M2-macrophage signature gene promoters were hypomethylated in fetal cells compared to maternal cells, indicating that DNA methylation is critical for maintaining immune tolerance at the feto-maternal interface [7]. Consistently, DNA methyltransferase 3b (DNMT3b) can regulate macrophage polarization (suppress the M1 phenotype and promote the M2 phenotype) via binding to the methylation region at PPARγ1 promoter in response to obesity-induced inflammation [8]. PPARγ can also act as a ligand-insensitive epigenomic regulator of chromatin structure that confers transcriptional memory by recruiting p300 and RAD21, which leads to more extracellular matrix related mRNA expression upon repeated treatment with IL4 [9]. As professional antigen-presenting cells, DCs are critical mediators of both innate and adaptive immune responses. Yet, it remains poorly understood how DC lineage differentiation and maturation are regulated at the epigenetic level. Recently, it was shown that dynamic DNA methylation changes during DC differentiation and maturation, and DNA demethylation around enhancers and binding sites of transcription factor occurred during DC maturation.

DC development and maturation are accompanied by loss of DNA methylation of CpG island in gene promoters including IL23R, IL10, CCR7, and CD59, which are attributed to downregulated DNA methyl-transferase expression (DNMT1, DNMT3A, and DNMT3B) [10]. 5-Azacytidine (5-aza), a hypomethylating agent, can inhibit the function of DNA methyltransferase 1. The costimulatory molecules CD40 and CD86 were significantly enhanced in 5-aza-treated mature DCs, which also promotes high expression of IFNγ and IL17A in activated T lymphocytes [11].

$CD4^+CD25^+Foxp3^+$ regulatory T cells (Treg cells) are critically required for maintaining self-tolerance and tissue homeostasis [12]. As a distinct subset, the first Treg cell-specific demethylated region in Foxp3 ‘conserved noncoding sequence 2’ (CNS2) was established in the thymus as compared to conventional T cells, which is critical for Treg identity. In addition, TCR stimulation can also trigger the induction of $Foxp3^+CD4^+$ Treg generation in the presence of TGFβ, with hypomethylation occurring at the CpG-rich TSDR (Treg-specific demethylation region) of $5'$ regulated region of Foxp3 to enhance its expression. Thus, the methylation signature of TSDR is demonstrated to efficiently control the Foxp3 transcriptional activity in vivo and in vitro [13]. As expected, Tregs from various peripheral tissues may share a common epigenetic signature. Of note, tissue-specific epigenetic reprogramming also occurs. For instance, methylation of Pparγ in adipose tissue Tregs was much lower as compared to relatively higher methylation in skin and LN Treg cells, indicating that a unique epigenetic landscape was acquired in distinct tissue to fulfill their specific function [14].

CD8 T cells play crucial roles in protective immunity to infection and cancer [15]. Upon antigen stimulation, naïve CD8 T cells undergo massive clonal expansion and differentiation into effector cells, with a small fraction of effector cells that will further differentiate into long-lived, self-renewing memory CD8 T cells. Understanding the molecular events governing their differentiation into long-lived memory cells following activation by antigen is a major question in immunology. In particular, a better understanding of the epigenetic mechanisms regulating the generation and persistence of long-lived CD8 memory T cells is highly relevant for the design of improved therapeutic vaccines. Of note, rather than retaining a naive epigenetic state, the memory precursor effector cells that give rise to memory cells did acquire de novo DNA methylation programs at naive-associated genes, and also acquired effector functions as the loci of classically defined effector molecules became demethylated. Consistently, T cell-specific deletion of the de novo methyltransferase Dnmt3a leads to reduced DNA methylation and more rapid reexpression of naïve or memory associated genes, indicating an accelerated memory T cell formation in the absence of DNMT3A [16]. Intriguingly, IFNγ and Granzyme B are hypermethylated at the loci of their promotors within memory T cells, but the rapid demethylation occurs to promote their expression upon recall response [17].

Despite great progress in cancer immunotherapy, a highly immunosuppressive/metabolically stressed tumor microenvironment remains a major barrier hindering effective antitumor immunity. In contrast to memory formation, under certain chronic infection circumstances or at the tumor microenvironment, CD8 T cells undergo exhaustion which was characterized by gradual loss of proliferative capacity and effector function, accompanied by upregulation of multiple co-inhibitory molecules. It remains largely unknown how epigenetic factors regulate T cell exhaustion or dysfunction. Recently, it was demonstrated that the epigenetic landscape is highly different from exhausted T cells as compared to effector or memory T cells [18]. The chromatin state dynamics underlying tumor-specific T cell dysfunction were defined to contain two phases with a plastic dysfunctional state and a fixed dysfunctional state that are resistant to reprogramming [19]. Interestingly, Dnmt3a mediated DNA de novo methylation is needed for acquisition of the exhaustion phenotype of CD8 T cells, and blocking these programs can enhance immune checkpoint blockade-mediated T cell rejuvenation [20].

RNA modification

As one of the most abundant mRNA modifications, N6-methyladenosine (m^6A) is involved in the regulation of multiple aspects of RNA biology, including RNA decay, splicing, and translation. The m^6A modification machinery consists of "writer," "reader," and "eraser." To explore the potential role of m^6A in regulating innate and adaptive immunity, genetic approaches were carried out to delete certain key regulators of m^6A machinery in defined immune subsets. For instance, specific ablation of Mettl3, the core component of the RNA methyltransferase complex, in dendritic cell (DC) leads to impaired phenotypic and functional maturation, with reduced capacity to stimulate T cell responses both in vitro and in vivo, indicating that Mettl3-mediated mRNA m^6A methylation is required for proper activation and function of DCs [21]. In addition, the deletion of either the m^6A 'writer' METTL3 or 'reader' YTHDF2 led to an increase in the induction of interferon-stimulated genes upon viral infection. Mechanistically, IFNβ was m^6A modified and was stabilized following the repression of METTL3 or YTHDF2 [22]. In addition, the nuclear DEAD-box (DDX) helicases member DDX46 could suppress type I interferon production upon viral infection via recruiting ALKBH5, an 'eraser' of the m^6A to demethylate many m6A-modified antiviral transcripts [23].

To explore the role of m^6A in T cell biology, Li et al. recently demonstrated that T cell-specific deficiency of the m^6A 'writer' enzyme Mettl3 leads to disrupted naïve T cell homeostasis, accompanied with enhanced inflammation in aged animals. Mettl3 deficient naïve T cells fail to expand in lymphopenic host upon adoptive transfer [24]. They further characterized the role of METTL3 in Treg cells and found that mice with METTL3 deficiency specifically in regulatory T cells develop severe autoimmune diseases after weaning [15], due to the largely impaired suppressive capacity in Treg cells in the absence of METTL3.

Histone modifications

Histone modifications, include methylation and acetylation, can alter chromatin structure and regulate target gene transcription in immune cells, which thus affects cell differentiation and functionality [25]. Histone methyltransferases (HMTs) are the enzymes that can add methyl groups to specific sites of histone proteins including H3K4 and H3K27, which can be methylated by MLL and EZH2, respectively [26]. A recent study shows that HMTs regulates monocytes/macrophages differentiation, and they found that the histone H3K27 region of FOXC1 is hypermethylated by EZH2. In addition, downregulation of FOXC1 represses KLF4, a key regulator of monocyte/macrophage differentiation [27]. Genes analysis of macrophage polarization shows that the H3K4-specific methyltransferase MLL is increased in M1 macrophage, which resulted in elevating H3K4me3 of the chemokine CXCL10. Thus, the polarization of the macrophages is dependent on the methylation including histone methylation. Histone methylation is also important to regulate differentiation and activation of DCs. Respiratory syncytial virus infection can interfere with innate inflammation and production of type I IFN. H3K4 demethylase KDM5b expression was upregulated in DCs following RSV infection, and histone demethylation resulted in repressed expression of antivirus genes and type I IFN [28]. Besides RSV infection, LPS can also induce several histone modifications in DCs, including histone methylation and histone acetylation. H3K9K14ac and H3K4me3 are accumulated at the transcriptional factors (TFs) binding sites involving the NF-κB and STAT1/2 family following LPS treatment in DCs [29].

Systems approach for deconvoluting immune cell composition

Of note, immune or tumor tissues represent a heterogeneous mixture of diverse cell types including epithelial cells, fibroblast, innate and adaptive immune cells, etc. In heterogeneous tissues, genome DNA sequences from the whole tissue could be extracted and used to measure the methylation level. In previous epigenome-wide association studies (EWAS), DNAm was shown to be the most feasible and relevant epigenetic modification [30]. Besides nucleotide modifications, alterations of the chromosome structure represent another layer of regulation regarding the development/differentiation of immune cells [31]. To date, many technologies including DNAse-seq, FAIRE-seq, MNase-seq, ATAC-seq, etc. have been developed to obtain the genome-wide chromatin states across different cell types [32, 33]. Of note, distinct trans-regulatory profiles were observed in different cell types with diverse functional differences [32]. Taking advantage of these cell-type-specific features, the DNAm data of a given tissue can be modeled as a linear summation of its cellular composition. Compared with traditional experimental methods, the immune cell composition in biological samples can be directly predicted from its omic data without dissecting them into single cells.

Deconvolution frameworks

Deconvolution methods have been developed to estimate the cellular composition from the methylation data, which can be mainly classified into two types: reference-based and reference-free methods (Table 1). For reference-based methods, predefined reference datasets representing the methylation profile of selected cell types and the corresponding computational framework were combined to formulate the computational model. Similar to the methods used to deconvolute the transcriptome data, the methylation level of each of the CpG sites is modeled as a linear summation for the methylation of its constituents, namely $\mathbf{A} = \mathbf{B}*\mathbf{X}$ (**A:** Methylation value for each of the CpG sites in tissue; **A:** Reference matrix for the DNAm in each cell type; **X:** Proportion for each cellular constituents). Utilizing a set of selected CpGs sites that can best discriminate different cell types, a training matrix $\mathbf{B}$ representing the methylation level of these selected CpGs sites across different cell types was constructed. With the input tissue methylation data $\mathbf{A}$, a machine learning algorithm was trying to find a $\mathbf{X}$ that can best minimize the residuals between $\mathbf{A}$ and $\mathbf{B}*\mathbf{X}$.

Reference-based models

Currently, several reference-based computational tools have been proposed. For instance, Houseman's CP/QP, CIBERSORT, robust multivariate linear regression, or robust partial correlations (RLR/RPC) have been proposed to infer the composition of major immune cell types from the human DNAm database (Fig. 1). With this DNAm data-based computation strategy, the relative proportions between different immune cells in blood, epithelial tissue, and cancer have been successfully estimated from whole-genome bisulfite sequencing data and RNA-seq data deposited in GEO and TCGA [37]. Similarly, these algorithms have been previously used in transcriptome data deconvolution. In this regard, by combining the deconvolution algorithm and the reference profile generated from the DNA methylation data of different immune cells, the abundance of tissue immune cells can be directly estimated from its methylation data. Using quadratic programming algorithms that have also been applied into the transcriptome data, a model was developed by Houseman et al. to infer proportions of eight immune cell types. Briefly, in CP/QP, inference proceeds via least-squares minimization but subject to the constraint that weights cannot be negative. Furthermore, the methylation data of blood samples collected from multiple disease settings including ovarian cancer, head and neck cancer, and Down syndrome have been tested with this model and consistent results were observed [34]. This model has already been packed into an R package called "Minfi" and released to the public through Bioconductor [39]. Next, using the support vector regression-based algorithm, which has been previously used in gene expression deconvolution, a model called "MethylCIBERSORT" was constructed to infer the tissue immune cell composition from the DNAm data of the whole tumor

Table 1 An overview of currently available deconvolution methods used in methylation data analysis.

Name	Language	Algorithm	Reference data	Cell type predicted	Tissue application	Reference
Housemen's CP/QP	R	Regression calibration approach algorithm	Yes	Six immune cell	Blood and tumor tissue	[34]
MethylCIBERSORT	R, Java	SVR	Yes	Fibroblasts and seven immune cell types	Tumor tissue	[35]
Epidish	R	Robust partial correlation	Yes	Seven immune cell types	Blood	[36]
HEpiDISH	R	robust partial correlation	Yes	Epithelial, fibroblast and, seven immune cell types	Tumor tissue	[37]
MeDeCom	R	Nonnegative matrix factorization	Reference-free		Blood and tissue	[38]

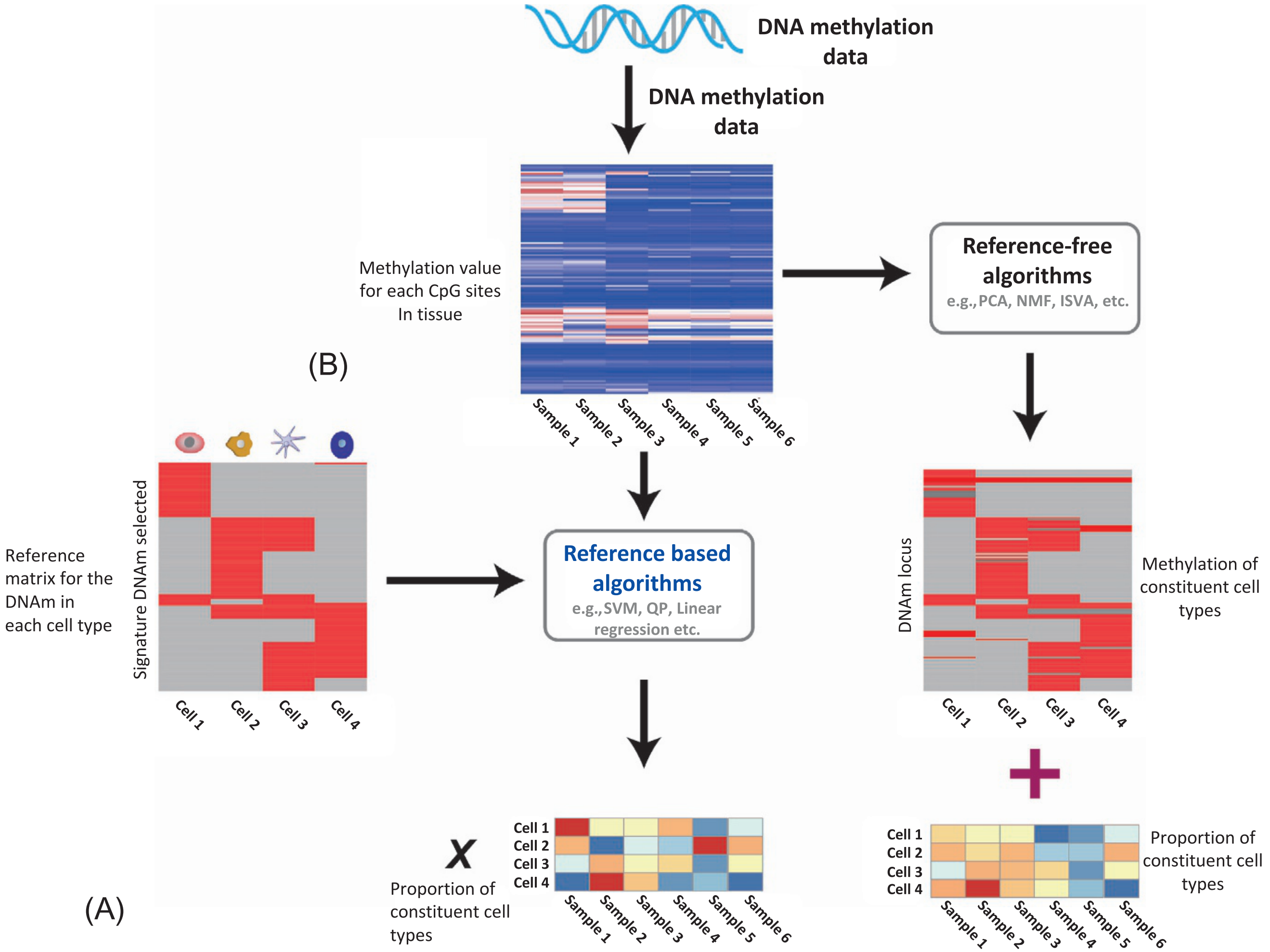

DNA methylation data
DNA methylation data
Methylation value for each CpG sites In tissue
Sample 1
Sample 2
Sample 3
Sample 4
Sample 5
Sample 6
Reference-free algorithms
e.g.,PCA, NMF, ISVA, etc.
(B)
Reference matrix for the DNAm in each cell type
Signature DNAm selected
Cell 1
Cell 2
Cell 3
Cell 4
Reference based algorithms
e.g.,SVM, QP, Linear regression etc.
DNAm locus
Methylation of constituent cell types
Cell 1
Cell 2
Cell 3
Cell 4
+
X
Proportion of constituent cell types
Cell 1
Cell 2
Cell 3
Cell 4
Sample 1
Sample 2
Sample 3
Sample 4
Sample 5
Sample 6
(A)
Cell 1
Cell 2
Cell 3
Cell 4
Proportion of constituent cell types
Sample 1
Sample 2
Sample 3
Sample 4
Sample 5
Sample 6

Fig. 1 Schematic representation of the reference-based and reference-free deconvolution tools to estimate the cellular composition of the tissue. Different strategies are defined according to the input data required and the constituent of output result. In reference-based tools, both methylation values of a given tissue and a reference matrix regarding the methylation of these selected loci in each cell type are required for the input, and a matrix for the cellular proportion in each sample was returned. For reference-free strategies, beyond the tissue methylation data, only limited knowledge for the cellular component was needed. After calculation, the matrix for the methylation of the possible constituent cell type and the correspondent cellular proportion was estimated.

tissue [35]. Tumor samples deposited in TCGA can be then classified into "immune hot" and "immune cold" according to the predicted immune cell composition. To examine which algorithm is more suitable, the performance of multiple algorithms including CP/QP, multivariate linear regression or partial correlations (LR), RLR/RPC and Support Vector Regressions (SVR) have been compared according to the RMSE and R^2 values [36]. Totally, the LR- and SVR-based models tend to be more robust to the inference of some realistic noise than CP/QP. Despite a relatively poor performance for CP/QP, application of this algorithm was still widely observed in multiple predictions [40–43]. Besides these previously mentioned on-step methods, the hierarchical strategy has also been applied to deconvolution model development. In HEpiDISH, the fraction of total immune cells and the relative proportion between seven immune cell types were separately estimated from tissue methylation data with RPC. By multiplying the relative proportion of each immune cell with the proportion of total immune cells, we can get the immune cell composition in that tissue [37].

Of note, the immune cells used for methylation profiling analysis were mainly collected from peripheral blood. Some tissue-resident immune cell subsets, like macrophages and DCs, are less presented in the computational model. In addition, macrophages and DCs usually account for only a small proportion in most nonimmune tissues. Furthermore, most types of immune cells can be further classified into different subtypes according to their differentiation status and functionality. For instance, CD4 T cells consisted of several functional subsets including Th1, Th2, Th9, Th17, Treg, etc. However, it was still challenging for us to obtain the methylation profiling data for all these immune subsets from the publicly available database. Therefore, building an appropriate calculation model without preparing a reference training matrix would further extend its application under certain physiological or pathological settings.

Reference-free models

In contrast to reference-based computational tools, the reference methylation profile was not required in reference-free models. Currently, several reference-free algorithms including principal component analysis (PCA), nonnegative matrix fraction (NMF), surrogate variable analysis (SVA), independent surrogate variable analysis (ISVA) have been specifically designed for methylation data deconvolution (Figure 1). According to the algorithm framework used in the model, reference-free tools including: ReFACTor, EWASher, RefFreeEWAS, fast-Lmm-eWasher [44] have been published.

Among these routine analysis strategies, PCA has been widely used to extract the variations existing in different samples and the first several principal components are considered to represent the major variance. Koestler et al. have shown that the methylation variation is correlated with the cell type composition [40]. Moreover, ReFACTor was constructed to evaluate the cellular heterogeneity of tissue. By performing a PCA on the profile of selected CpG sites across several samples, the underlying cellular composition can be obtained [45]. Further, combining the major principal components representing cell composition information captured with ReFACTor with a bayesian strategy, a semi-supervised method BayesCCE was constructed [46]. Specifically, except for the cell composition, components for each cell type can also be generated and used for later analysis. Similar to reference-based computational tools, the methylation data for each sample can still be regarded as the product between a matrix of training data and a vector of the cellular proportion. With a heuristic for constrained matrix factorization using quadratic programming, both the cellular proportion and the methylation profile of each component can be estimated from the methylation data [47]. Combined with the transcriptome data, the transcription profile for each cell type can also be successfully inferred.

Perspectives

Using these deconvolution tools, it is convenient for us to obtain the immune cell compositions from the transcriptome data of a given tissue. This knowledge may greatly advance our understanding of the complex regulation within the immune system. Despite the fact that certain success was achieved as mentioned above, there are still many key questions remaining to be solved to increase the performance rate. Firstly, these computational models should be extended to more immune cell subtypes. Currently, the number of immune cell subsets predicted by these tools was ranging from 4 to 10. For instance, DCs and mast cells are not included in previously mentioned computation models due to limited resources of available methylation data. Additionally, immune cells can change their phenotypic marker expression pattern, and many types of immune cells are rather plastic to further differentiation into other subsets. It represents an important step to further define those different immune subtypes according to their differentiation status or functional roles. Yet, the methylation data for these cell types including differentiation and maturation of dendritic cell [10], exhausting T cell [18], have been reported. Apart from that, an atlas for the methylation profile of 25 tissues and cell types have also been categorized into a database [48]. Thus, by adding the methylation information of these immune cells into our previous training signature matrix, it would be helpful to get more comprehensive information about the constitution of immune cells within defined tissues.

Since the training data consisted of the methylation level of those selected signature loci across different immune cells, the selected signature DNAm loci are therefore important for model accuracy and calculation speed. Yet, the mean difference in DNAm between each cell type was the major parameter for consideration during the signature locus selection process. However, as the potential collinear relationship between some DNAm locus, certain bias will be produced in some situations when calculating with all these selected loci. With a PCA step, these selected features can be converted into some principal components [49]. Further, to reduce the bias, a conditional number minimization algorithm used in the expression data deconvolution model can also be applied here to improve matrix stability [50].

Since the collinear relationship existed among cell types when the expression profile was similar, it was difficult to well distinguish each other with the previously described one-step deconvolution models. Inspired by the gating strategy used in flow cytometry data analysis, deconvolution with a hierarchical model will be helpful to accurately quantify immune subtypes. Specifically, the proportion of the major immune cell types and the corresponding subtypes was estimated in different steps. A straightforward strategy is to predict the proportion and the methylation profiling of a defined immune cell type from the total methylation profile data of the heterogeneous tissues at the first step. Then, with the methylation profile of major immune cells, the relative proportion between these cell subtypes can be subsequently inferred. Recently, a nonnegative least-squares regression (NNLS) based method has been successfully implemented to determine the proportion and a representative expression profile for each cellular component from the transcriptome data [51]. Further integration of this framework into the deconvolution of methylation data will foster the development of a hierarchical model. Besides the methylation profiles of immune cells, the training signature matrix used in different hierarchical steps was also an important element. During the differentiation of different immune lineages, the methylation and demethylation pattern of CpG loci can be selectively regulated to tightly control the transcription of downstream genes. Similar to the gene expression profiles, cell-type specificity can also be characterized by their methylation profile. For instance, the demethylation of CD3 locus was observed in all T cells and may represent a T cell determining factor. Likewise, the promoter of IFNγ was demethylated in Th1 cell and Th17 cell and methylated in Th2 cell [1]. According to the immune cells to be estimated, the signature CpG islands used in the training matrix will also be different. Specifically, the major cell type determining locus was taken into the training matrix for the first deconvolution process while the cell subtype-specific methylation region should be only included in the second computation step.

As expected, the methylation profile of immune cells can also be affected by the tissue microenvironment. To better functionally adapt to the surrounding environment, immune cells localized within different tissues may exhibit strikingly distinct methylation profiles. For instance, a distinct enhancer landscape was identified in tissue-resident macrophages and this tissue specific feature can be shaped by tissue microenvironment [52]. Similarly, with respect to the methylation profile of tissue-infiltrated Treg cells, nearly 11,000 differentially methylated regions have been identified. Therefore, it will be helpful to improve the performance when taking this tissue-specific information into consideration. With the development of single-cell sequencing technology, genome-wide methylation patterns can be easily characterized at the single-cell level. With scRNA-Seq, a tissue-specific deconvolution strategy has already been successfully implemented into the expression data-based model to improve model performance [53]. Similarly, model accuracy can also be improved if this strategy could be integrated into the methylation based deconvolution model.

Finally, beyond methylation profiling, other specific epigenetic features attributed to a defined cell type can also be measured with other omics technology. Hence, the accuracy of modeling may be further improved when multi-omic data including transcriptome, chromatin structure, and conformation accessibility was integrated. In a previous methylation data deconvolution model, a modest improvement was observed when combing the DNAse Hypersensitive Site (DHS) information with the DNA methylation [36]. In particular, it is possible that multi-omic data can be simultaneously studied at the single-cell level. For instance, combing the TCR-encoding gene sequencing with transposase-accessible chromatin with sequencing (ATAC-seq), cis and trans regulators of different T cell subtypes can be successfully identified [33]. Therefore, important information regarding cell type-specific features can be obtained from the multi-omic data. Altogether, with the potential strategies discussed above, this methylation-based immune cell deconvolution model can be widely applied to other epigenetic analysis.

References

[1] Suarez-Alvarez B, et al. DNA methylation: a promising landscape for immune system-related diseases. Trends Genet 2012;28(10):506–14.
[2] Jones PA, et al. Epigenetic therapy in immune-oncology. Nat Rev Cancer 2019;19(3):151–61.
[3] Wessels I, et al. Changes in chromatin structure and methylation of the human interleukin-1beta gene during monopoiesis. Immunology 2010;130(3):410–7.
[4] Lorente-Sorolla C, et al. Inflammatory cytokines and organ dysfunction associate with the aberrant DNA methylome of monocytes in sepsis. Genome Med 2019;11(1):66.
[5] Bullwinkel J, et al. Epigenotype switching at the CD14 and CD209 genes during differentiation of human monocytes to dendritic cells. Epigenetics 2011;6(1):45–51.
[6] Takeuch O, Akira S. Epigenetic control of macrophage polarization. Eur J Immunol 2011;41(9):2490–3.
[7] Kim SY, et al. Methylome of fetal and maternal monocytes and macrophages at the feto-maternal interface. Am J Reprod Immunol 2012;68(1):8–27.
[8] Yang X, et al. Epigenetic regulation of macrophage polarization by DNA methyltransferase 3b. Mol Endocrinol 2014;28(4):565–74.
[9] Daniel B, et al. The nuclear receptor PPARgamma controls progressive macrophage polarization as a ligand-insensitive epigenomic ratchet of transcriptional memory. Immunity 2018;49(4):615–26. e6.
[10] Zhang X, et al. DNA methylation dynamics during ex vivo differentiation and maturation of human dendritic cells. Epigenetics Chromatin 2014;7:21.
[11] Frikeche J, et al. Impact of the hypomethylating agent 5-azacytidine on dendritic cells function. Exp Hematol 2011;39(11):1056–63.
[12] Fontenot JD, et al. Foxp3 programs the development and function of CD4 + CD25 + regulatory T cells. Nat Immunol 2003;4(4):330–6.
[13] Polansky JK, et al. DNA methylation controls Foxp3 gene expression. Eur J Immunol 2008;38(6):1654–63.
[14] Delacher M, et al. Genome-wide DNA-methylation landscape defines specialization of regulatory T cells in tissues. Nat Immunol 2017;18(10):1160–72.

[15] Tong J, et al. m(6)A mRNA methylation sustains Treg suppressive functions. Cell Res 2018;28(2):253–6.
[16] Youngblood B, et al. Effector CD8 T cells dedifferentiate into long-lived memory cells. Nature 2017;552(7685):404–9.
[17] Fitzpatrick DR, et al. Cutting edge: stable epigenetic inheritance of regional IFN-gamma promoter demethylation in CD44highCD8+ T lymphocytes. J Immunol 1999;162(9):5053–7.
[18] Sen DR, et al. The epigenetic landscape of T cell exhaustion. Science 2016;354(6316):1165–9.
[19] Philip M, et al. Chromatin states define tumour-specific T cell dysfunction and reprogramming. Nature 2017;545(7655):452–6.
[20] Ghoneim HE, et al. De novo epigenetic programs inhibit PD-1 blockade-mediated T cell rejuvenation. Cell 2017;170(1):142–57. e19.
[21] Wang H, et al. Mettl3-mediated mRNA m(6)A methylation promotes dendritic cell activation. Nat Commun 2019;10(1):1898.
[22] Winkler R, et al. m(6)A modification controls the innate immune response to infection by targeting type I interferons. Nat Immunol 2019;20(2):173–82.
[23] Zheng Q, et al. The RNA helicase DDX46 inhibits innate immunity by entrapping m(6)A-demethylated antiviral transcripts in the nucleus. Nat Immunol 2017;18(10):1094–103.
[24] Li HB, et al. m(6)A mRNA methylation controls T cell homeostasis by targeting the IL-7/STAT5/SOCS pathways. Nature 2017;548(7667):338–42.
[25] Zhang Q, Cao X. Epigenetic regulation of the innate immune response to infection. Nat Rev Immunol 2019.
[26] Bermick J, et al. Histone methylation is critical in monocyte to macrophage differentiation. FEBS J 2017;284(9):1306–8.
[27] Somerville TD, et al. Frequent derepression of the mesenchymal transcription factor gene FOXC1 in acute myeloid leukemia. Cancer Cell 2015;28(3):329–42.
[28] Ptaschinski C, et al. RSV-induced H3K4 demethylase KDM5B leads to regulation of dendritic cell-derived innate cytokines and exacerbates pathogenesis in vivo. PLoS Pathog 2015;11(6):e1004978.
[29] Vandenbon A, et al. Waves of chromatin modifications in mouse dendritic cells in response to LPS stimulation. Genome Biol 2018;19(1):138.
[30] Baron U, et al. DNA methylation analysis as a tool for cell typing. Epigenetics 2006;1(1):55–60.
[31] Winter DR, Amit I. The role of chromatin dynamics in immune cell development. Immunol Rev 2014;261(1):9–22.
[32] Grbesa I, et al. Mapping genome-wide accessible chromatin in primary human T lymphocytes by ATAC-Seq. J Vis Exp 2017;129.
[33] Satpathy AT, et al. Transcript-indexed ATAC-seq for precision immune profiling. Nat Med 2018; 24(5):580–90.
[34] Houseman EA, et al. DNA methylation arrays as surrogate measures of cell mixture distribution. BMC Bioinform 2012;13:86.
[35] Chakravarthy A, et al. Pan-cancer deconvolution of tumour composition using DNA methylation. Nat Commun 2018;9(1):3220.
[36] Teschendorff AE, et al. A comparison of reference-based algorithms for correcting cell-type heterogeneity in epigenome-wide association studies. BMC Bioinform 2017;18(1):105.
[37] Zheng SC, et al. A novel cell-type deconvolution algorithm reveals substantial contamination by immune cells in saliva, buccal and cervix. Epigenomics 2018;10(7):925–40.
[38] Lutsik P, et al. MeDeCom: discovery and quantification of latent components of heterogeneous methylomes. Genome Biol 2017;18(1):55.
[39] Aryee MJ, et al. Minfi: a flexible and comprehensive bioconductor package for the analysis of Infinium DNA methylation microarrays. Bioinformatics 2014;30(10):1363–9.
[40] Koestler DC, et al. Blood-based profiles of DNA methylation predict the underlying distribution of cell types: a validation analysis. Epigenetics 2013;8(8):816–26.
[41] Cardenas A, et al. Validation of a DNA methylation reference panel for the estimation of nucleated cells types in cord blood. Epigenetics 2016;11(11):773–9.
[42] Kong Y, et al. Insights from deconvolution of cell subtype proportions enhance the interpretation of functional genomic data. PLoS One 2019;14(4):e0215987.

[43] Wen Y, et al. Cell subpopulation deconvolution reveals breast cancer heterogeneity based on DNA methylation signature. Brief Bioinform 2017;18(3):426–40.
[44] Zou J, et al. Epigenome-wide association studies without the need for cell-type composition. Nat Methods 2014;11(3):309–11.
[45] Rahmani E, et al. Sparse PCA corrects for cell type heterogeneity in epigenome-wide association studies. Nat Methods 2016;13(5):443–5.
[46] Rahmani E, et al. BayesCCE: a Bayesian framework for estimating cell-type composition from DNA methylation without the need for methylation reference. Genome Biol 2018;19(1):141.
[47] Onuchic V, et al. Epigenomic deconvolution of breast tumors reveals metabolic coupling between constituent cell types. Cell Rep 2016;17(8):2075–86.
[48] Moss J, et al. Comprehensive human cell-type methylation atlas reveals origins of circulating cell-free DNA in health and disease. Nat Commun 2018;9(1):5068.
[49] Li J, et al. Machine learning methods for predicting human-adaptative influenza A viruses based on viral nucleotide compositions. Mol Biol Evol 2019;msz276.
[50] Newman AM, et al. Robust enumeration of cell subsets from tissue expression profiles. Nat Methods 2015;12(5):453–7.
[51] Newman AM, et al. Determining cell type abundance and expression from bulk tissues with digital cytometry. Nat Biotechnol 2019;37(7):773–82.
[52] Lavin Y, et al. Tissue-resident macrophage enhancer landscapes are shaped by the local microenvironment. Cell 2014;159(6):1312–26.
[53] Chen Z, et al. Tissue-specific deconvolution of immune cell composition by integrating bulk and single-cell transcriptomes. Bioinformatics 2020;36(3):819–27.

CHAPTER 11

Epigenetic deregulation of immune cells in autoimmune and autoinflammatory diseases

Javier Rodríguez-Ubreva*, **Tianlu Li***, **Esteban Ballestar**
Epigenetics and Immune Disease Group, Josep Carreras Research Institute (IJC), Barcelona, Spain

Contents

Relevance of epigenetics for immune deregulation in autoimmune/autoinflammatory disorders 253
Epigenetic dysregulation in autoimmune diseases 254
  Epigenetic defects of immune cells in rheumatoid arthritis 254
  Epigenetic defects of immune cells in psoriasis 256
Epigenetic dysregulation in autoinflammatory diseases 258
  Familial Mediterranean fever 258
  Cryopyrin-associated periodic syndromes 259
Epigenetic biomarkers in autoimmunity 260
Targeting epigenetic defects 262
References 263

Relevance of epigenetics for immune deregulation in autoimmune/autoinflammatory disorders

Self-inflammatory alterations have been classically organized into autoimmune diseases, when they imply altered adaptive immune responses, and autoinflammatory diseases, when the innate immune compartment is involved. However, it appears clearer in the last decade that self-inflammatory conditions include a plethora of intermediate states in which autoinflammation and autoimmunity are the extremes of the same inflammatory spectrum. In the attempt to unravel the molecular mechanisms involved in immune-altered responses in autoimmune/inflammatory diseases, their etiology has been associated with genetic susceptibility in combination with epigenetic alterations deriving from environmental factors. In this context, epigenetic deregulation has emerged as a contributor in the pathophysiology of the disease as well as an element that can influence the outcome of the disease and treatments efficacy.

Epigenetic modifications influence gene expression and modulate cellular functions without modifying the genomic sequence, and involve two major players: DNA

---

* These authors contributed equally to this chapter.

Epigenetics of the Immune System
https://doi.org/10.1016/B978-0-12-817964-2.00011-3

© 2020 Elsevier Inc.
All rights reserved.

methylation and histone modifications. DNA methylation has been described to have a crucial role in transcriptional regulation and in cell fate decisions as well as cell functions. It consists the addition of a methyl group to the 5′ position of the pyrimidine ring of certain cytosines adjacent to guanines (named CpG sites). This chemical modification is catalyzed by DNA methyltransferases (DNMTs), including DNMT1, DNMT3A, and DNMT3B. On the contrary, TET family proteins (TET1, TET2, and TET3), methylcytosine dioxygenases that act in coordination with thymine DNA glycosylase, are necessary for active DNA demethylation [1]. TET enzymes catalyze the generation of oxidized versions of methylcytosine that not only are intermediates in the demethylation process but also functional nucleotides on their own. Another important epigenetic mechanism involves posttranslational modification of histone amino acid residues such as Lys acetylation, Lys and Arg methylation, Ser and Thr phosphorylation, and Lys ubitiquination. These chemical modifications modulate the interactions between DNA and specific nuclear factors as well as chromatin structure [2].

Epigenetic dysregulation in autoimmune diseases

The main feature of autoimmune diseases is the pathological activation of the host immune cells, which mount immune responses against self-antigens, generating autoantibodies. Among these autoimmune diseases, two main groups can be distinguished based on their clinical manifestations: systemic and organ-specific autoimmune disorders. The term systemic autoimmune rheumatic diseases (SARDs) are used to refer to chronic inflammatory diseases including rheumatoid arthritis (RA) and systemic lupus erythematosus (SLE) among others. On the contrary, other autoimmune disorders only affect specific organs, such as psoriasis and diabetes mellitus. In this chapter, we will summarize the epigenetic alterations affecting immune cells in the context of RA and psoriasis as two examples of that autoimmune spectrum.

Epigenetic defects of immune cells in rheumatoid arthritis

Rheumatoid arthritis (RA) is a systemic autoimmune disease of connective tissue, characterized by progressive joint damage and systemic organic alterations. The etiology of this disease remains unknown but it appears that genetic predisposition, together with environmental factors, are involved in pathogenesis and disease progression. The pathological development of RA consists of an autoimmune inflammation of the synovial membrane of joints, involving synovial cells proliferation and pannus formation, an abnormal and aggressive layer of granulation tissue that induces articular cartilage erosion and bones destruction. This joint dysfunction promotes the infiltration of immune cells to the synovial environment. In this context, T cells release different pro-inflammatory molecules and B cells produce autoantibodies, such as rheumatoid factor (RF) and anti-cyclic citrullinated peptide (anti-CCP). Differences in anti-CCP and RF levels, and variability of response to treatments among other factors, indicate different pathophysiological mechanisms during RA progression [3, 4].

Peripheral blood or peripheral blood mononuclear cells (PBMCs) consist a mixture of different immune cell types, including monocytes, natural killer, $CD4^+$ and $CD8^+$ T cells and B cells among others, whose proportions may change under pathological conditions. This enormous heterogeneity might be masking some of the altered methylation patterns associated to RA that occur only in certain cell subpopulations at specific steps of disease progression. For that reason, different approaches using whole blood and PBMC samples do not appear to be optimal to identify the underlying epigenetic alterations in RA in specific immune cell compartments. Nevertheless, the analysis of DNA methylation using PBMCs has been successful as clinical markers of RA predisposition and has been used to anticipate the responses to certain treatments. For instance, Liu et al. performed an epigenome-wide association study using whole blood and identified 10 differently methylated CpGs that constitute two clusters containing genes of the major histocompatibility complex (MHC), which indeed are a risk factor in RA [5]. In this line of study, van Steenberg et al. demonstrated that the methylation of a specific CpG site within the major histocompatibility complex region was also significantly associated with RA [6]. Another study with PBMCs from RA patients indicated that hypermethylation of four CpG sites located at exon 7 of *LRPAP1* gene was associated with the lack of response to anti-TNF therapy (etanercept) in comparison with responders [7].

In addition, in the last years, several studies have demonstrated the existence of epigenetic alterations in specific immune cells from RA patients. For instance, de Andres and colleagues observed an aberrant global hypomethylation of monocytes and T cells from RA individuals that is reversed after methotrexate (MTX) treatment [8]. In Treg cells from RA patients, Cribbs et al. showed that the hypermethylation of the promoter of *CTLA4* gene blocks the binding of the transcription factor NFATc2. This leads to *CTLA4* downregulation and, as a consequence, Treg cells are unable to induce the expression of the enzyme indoleamine 2,3-dioxygenase (IDO), which abolishes the immunosuppressive function of these cells.

In the context of RA B cells, a study identified differentially methylated CpGs in several genes. Interestingly, these differences were reproduced in a cohort of patients with SLE, suggesting the existence of similar drivers for epigenetic alterations in these two systemic autoimmune diseases [9].

Regarding the myeloid compartment, it has been described that the disease activity of RA patients can be directly stratified from DNA methylation patterns of peripheral blood monocytes. Interestingly, monocytes from the same RA individual change according to her/his disease activity status, indicating the plasticity of these immune cells. The authors demonstrated that there is a link between RA disease activity and the monocyte methylome through the action of inflammation-associated cytokines such as IFN-γ, IFN-α, and TNF-α. Interestingly, they observed that inflammation-associated methylation patterns are also present in other autoinflammatory conditions such as multiple myeloma, again suggesting the existence of common drivers in the establishment of altered methylomes in different autoimmune diseases [10].

Several studies have described alterations in histone posttranslational patterns and histone-modifying enzymes in RA. For instance, PBMCs isolated from RA patients exhibit significantly higher histone deacetylase (HDAC) activity compared to PBMCs from healthy individuals. The use of the specific HDAC inhibitor MI192 blocks TNF and IL-6 production in RA PBMCs but not healthy PBMCs [11]. Accordingly, both the activity and the levels of HDAC3 are reduced in PBMCs from RA patients, which is accompanied with enhanced histone acetyl transferase (HAT) activity [12]. In addition, alterations in histone variants have also been described in the context of RA. Histone H2A.Z and histone H3.3 have been shown to be more abundant in PBMCs from RA patients. Since these two histone variants have a prominent role in nucleosome positioning, their alterations could affect chromatin accessibility and the establishment of proper transcriptional programs. Interestingly, the authors also found a correlation between the levels of the histone H2A.Z and the disease activity score in RA patients [13].

Epigenetic defects of immune cells in psoriasis

Psoriasis is a chronic autoimmune/inflammatory disease characterized by skin manifestations. It has been estimated that the incidence of the disease in the developed countries is 1%–4% and it is frequently linked to disability and a significant decrease in the quality of life [14, 15]. Psoriasis can coexist with other pathologies such as psoriatic arthritis, ankylosing arthritis, and inflammatory bowel diseases. Indeed, approximately 20% of the patients with psoriasis are affected by psoriatic arthritis [16]. This disease is triggered by the combination of both genetic and environmental elements (connected with the epigenetic machinery), and it displays features of a mixed-pattern disease, where the activation of either innate (autoinflammatory) or adaptive (autoimmune) immune responses predominates each disease stage [17]. In the first stages of the disease, but also during flares, effector cells of the innate immune system, such as macrophages, neutrophils, and mast cells, are involved in cutaneous infiltration, and pro-inflammatory molecules such as IL-1β, TNF-α, and IFN-γ are essential to drive the inflammatory environment [18]. In this context, an IL-1-Th17-dominated autoinflammation is involved in the initiation of psoriasis. Indeed, targeting IL-17 using humanized monoclonal antibodies ameliorates clinical symptoms of psoriasis [19]. Although the initial inflammation is mediated by Th17 cells and downstream responses, IFNγ-producing Th1 cells drive subsequent inflammation in stable psoriasis plaques [18, 20]. The altered expression of these cytokines, in particular IL-17, IFN-γ, or TNF-α, might induce the pathological proliferation of keratinocytes. In addition, stimulated keratinocytes produce other molecules that perpetuate inflammation as a consequence of the activation and the recruitment of more immune cells to psoriatic lesions [21]. Furthermore, regulatory T cells have a central role in self-tolerance since they suppress the activation and proliferation of effector cells through the release of IL-10 and, consequently, controlling the inflammatory events in the organism [22]. Remarkably, both the regulatory activity of these cells, as well as their number in psoriatic lesions, are reduced [23, 24].

Over the past few years, several groups have focused their attention in analyzing epigenetic defects in keratinocytes and cutaneous lesions. However, very little is known about epigenetic aberrancies associated to psoriasis pathogenesis and development within the immune compartment. In this regard, initial studies showed a significant increase in global DNA methylation, as well as DNMT1 expression, in psoriatic PBMCs compared with healthy controls [25]. In a pioneer study using CD4$^+$ and CD8$^+$ T cells from monozygotic discordant twins for psoriasis, Gervin et al. identified differential DNA methylation in CD4$^+$ T cells of affected compared to unaffected twins in several psoriasis-related genes and in immune-related cytokine and chemokine genes. Interestingly, these methylation differences correlated with differential gene expression, suggesting that epigenetic alterations potentially contribute to disease development [26]. In this line, another study with CD4$^+$ T cells showed hypomethylation in peri-centromeric regions and hypermethylation in the promoter of immune-related genes in the X-chromosome of psoriasic patients compared to atopic dermatitis patients and healthy controls [27]. Additionally, treatment with the methylation inhibitor 5-azacytidine *in vitro* induced the reexpression of some hypermethylated genes in psoriasic CD4$^+$ T cells [28]. Another study showed that in CD4$^-$CD8$^-$CD3$^+$TCR+ double negative (DN) T cells of psoriasis patients, a distal regulatory element of the gene that encodes IFN-γ displays hypomethylation. Indeed, the authors showed that DN effector T cells are able to infiltrate the swollen tissues promoting inflammation and damage. Given the strong Th1 inflammatory responses present in the plaques of psoriasis patients, the elevated IFN-γ expression in DN T cells, a consequence of aberrant hypomethylation, might be relevant for the pathogenesis of psoriasis [29].

There are limited studies regarding histone modification patterns established during psoriasis development in the different cells of the immune system. Ovejero-Benito et al. demonstrated a global loss of acetylation in histones H3 and H4 in PBMCs isolated from psoriasis patients, as well as an increase in H3K4 methylation. Unfortunately, the analysis did not specify the loci where these changes are taking place nor the specific modifications (mono-, di-, or tri-methylation). Nevertheless, the authors observed an interesting data, in which although H3K27 methylation did not change when comparing healthy controls and untreated psoriasis patients, the levels of that histone modification were increased only in those patients that responded to treatments with biological drugs. Additionally, by excluding patients with coexisting arthritis, the authors also observed significant changes in H3K4 methylation between responders and nonresponders to biological drugs [30]. Along this line of research, another study indicated that histone H4 was hypoacetylated in PBMCs from patients with psoriasis. Indeed, there was an inverse correlation between the levels of histone H4 acetylation and disease activity. Furthermore, the gene expression levels of several histone remodelers, such as P300, CBP, and SIRT1, were reduced, whereas HDAC1, SUV39H1, and EZH2 were significantly increased in patients with psoriasis [31].

In the last years, research has been focusing on the contribution of the adaptive immune responses in the onset and development of psoriatic inflammatory conditions. However, the relevance of the innate immune responses has been underestimated despite the existence of several studies showing how polymorphisms in molecules of the inflammasome and NFκB signaling pathway increment disease risk [32, 33]. In this regard, to date, there are no specific studies that highlight the relevance of epigenetic deregulation in cells of the innate immune system, such as monocytes/macrophages, despite the contribution of the myeloid compartment in psoriasis [34, 35]. Understanding the complex interactions between all the components of the immune system in skin homeostasis and how epigenetic alterations might cause imbalance in the correct function of these cells during psoriasis pathogenesis can help to develop new and efficient strategies for treatment and prevention.

Epigenetic dysregulation in autoinflammatory diseases

Autoinflammatory disorders are a group of diseases that are essentially characterized as either systemic or organ-specific inflammation that manifests as recurrent fevers, high acute phase responses and a tendency to develop inflammation in joints, skin, and other organs. These disorders lack the hallmarks of other autoimmune diseases and are predominantly defined by the absence of autoreactive lymphocytes and high-titer autoantibodies. Historically, autoinflammatory conditions were identified by strong family histories and characteristic clinical manifestations with specific mutations in inflammation-related genes that follow the classical Mendelian mode of inheritance. Hence, these disorders are referred to as monogenic diseases. However, a wide range of clinical symptoms presented by family members with the same underlying mutation and the existence of individuals that display related symptoms, even in the absence of genetic mutations, suggest that other mechanisms, such as epigenetic dysregulation, may play a major role in the etiology of these disorders.

Familial Mediterranean fever

The best-known example of autoinflammatory diseases is Familial Mediterranean fever (FMF), which is characterized by recurrent fevers and inflammation in skin, joints, and peritoneum. Originally, FMF patients were identified to carry mutations in both alleles of *MEFV* gene, which encodes pyrin, a pattern-recognition receptor expressed on the surface of innate immune cells [36]. Pyrin modulates caspase-1 activity through its interaction with the adaptor protein ASC (apoptosis-associated speck-like protein with a caspase-recruitment domain), which in turn activates IL-1β [37]. Recently, several studies have shown that the genetic link between *MEFV* mutations and disease onset is not so straightforward [38], in which as much as more than half of the patients were found to not harbor any mutations [39]. Furthermore, individuals carrying the same mutations may also present a spectrum of symptoms that ranges in severity [40], and more surprisingly, healthy individuals have been described to carry mutations in the *MEFV* gene [41]. Therefore, it is rational to envision that other factors, such as epigenetics, can contribute to disease development. Aberrant DNA methylation of the *MEFV* gene in FMF patients was first identified by Kirectepe et al. [42]. An increase in DNA methylation of the second exon of *MEFV* was found to correlate with a decrease in its gene expression in peripheral leukocytes of FMF patients. These results were then validated by Erdem et al. using an *in vitro* system, where the authors show that aberrant DNA methylation of the *MEFV* gene lead to the expression of an aberrant spliced form of pyrin, which is elevated in FMF patients [43, 44]. Altogether, these evidence point to DNA methylation as a possible culprit in contributing to the pathology of FMF by altering the correct expression of pyrin.

Cryopyrin-associated periodic syndromes

Cryopyrin-associated periodic syndromes (CAPS) envelope a spectrum of several autoinflammatory disorders of distinct disease phenotypes. These include three disorders of increasing severity: familial cold autoinflammatory syndrome (FCAS), Muckle-Wells syndrome, and neonatal onset multisystem inflammatory disease (NOMID). One common genetic feature of these syndromes is the presence of gain-of-function mutations in the *NLRP3* gene, which encodes the NALP3 protein, otherwise known as cryopyrin. NALP3 belongs to a family of pattern-recognition receptors, which are activated by binding danger-associated molecular pattern molecules (DAMPs). Mutations in the *NLRP3* gene lead to increased aberrant activation of NALP3, which recruits several proteins, including caspase 1, to form the inflammasome [45]. Activated caspase 1 subsequently cleaves the inhibitory domains of IL-1β and IL-18, thus mediating their activation and release [46]. Both IL-1β and IL-18 are potent pro-inflammatory cytokines that promote the recruitment and activation of both innate and adaptive immune cells, hence causing the underlying systemic inflammatory observed in CAPS patients. Nevertheless, CAPS patients present a wide range of symptoms and the type of mutations does not correlate well with disease severity [47], which altogether suggest the involvement of other mechanisms. Possible epigenetic alterations were first described in patients with NOMID, where several genes encoding histone modification enzymes, such as *MLL* and *SIRT1*, and enzymes that modulate DNA methylation, including *TET2* and *DNMT3L*, were found to be differentially expressed between lesional and non-lesional skin [48]. A subsequent study by Vento-Tormo and colleagues [49] convincingly showed that monocytes isolated from CAPS patients underwent aberrant demethylation of inflammasome genes following IL-1β stimulation, which corresponded to their over-expression. This aberrant demethylation was reversed in patients receiving anti-IL-1β therapy, which not only indicates the importance of DNA methylation alterations in disease progression, but also reveals DNA methylation as a target of effective treatment strategy [49]. Furthermore, degree of DNA methylation of inflammasome genes were not found to correlate with the type of mutations in the *NLRP3* gene harbored by the CAPS patients, which again highlights epigenetic alterations as independent events that play pivotal roles in defining disease phenotype.

In conclusion, although genetic aspects of monogenic autoinflammatory diseases are well defined, it does not explain intraindividual variation in regards to disease onset, severity and response to treatment. Epigenetic alterations may partially provide the answers to better understand these diseases, and can serve as a promising target for personalized therapeutic intervention and/or biomarker identification in the near future.

Epigenetic biomarkers in autoimmunity

The study of epigenetics in the context of autoimmunity is not only useful for understanding the deregulation of mechanistic components of each disease, but also can be a potent tool for diagnosis, to estimate disease activity, and predict response to therapy. Heterogeneity of symptoms and disease progression, as well as varied response to therapy highlight the urgent need for personalized approaches in regards to diagnosis and treatment of autoimmune and autoinflammatory diseases. Recent technological advances in omics data analyses have allowed the exponential generation of a large quantity of data; however, more research is still required to effectively process, store, and extract this information with the final purpose of finding efficacious treatment for each patient. Examples are presented in Table 1. One such example of using omics data for diagnosis is a recent study by Chen and colleagues, in which the authors identified aberrant hypomethylation in several type I IFN-related genes in CD8$^+$ and CD4$^+$ T cells isolated from patients diagnosed with Graves' disease (GD), RA, SLE, and systemic sclerosis (SSc) [50]. Furthermore, the level of methylation detected at these genes appeared to be specific to each disease, which showed both high prediction ability and diagnostic potential [50]. Clinical manifestations of autoimmune diseases translate to several measurable parameters, one of which is disease activity. Several studies have recently been able to identify a direct link between aberrant epigenetic modifications and disease activity. First, in one study, hypomethylation of *CYP2E1* in monocytes and CD4$^+$ T cells isolated from RA patients was associated with high disease activity, and *DUSP22* hypomethylation in CD19$^+$ B cells

Table 1 Epigenetic biomarkers in inflammatory rheumatic diseases.

Disease	Sample type	Epigenetic alteration	Comments	Ref
GD, RA, SLE, SSc	$CD3^+$ and $CD4^+$ T cells	DNA hypomethylation of IFN-related genes	Degree of DNA hypomethylation display high specificity to each disease	[50]
RA	$CD4^+$ T cells and monocytes	DNA hypomethylation of *CYP2E1* promoter	Hypomethylation correlates with high disease activity	[51]
	$CD19^+$ B cells	DNA hypomethylation of *DUSP22* promoter	Hypomethylation correlates with the diagnosis of erosive RA	
RA	Monocytes	Alterations in DNA methylation of monocytes	Degree of DNA methylation correlates with DAS28 disease activity	[10]
Psoriasis	PBMCs	Aberrant histone H4 acetylation	Inverse correlation between histone H4 acetylation and disease activity	[31]
RA	$CD3^+$ T cells	21 differentially methylated CpGs	DNA methylation at diagnosis predict responsiveness to DMARDs	[52]
RA	Whole blood	DNA methylation of two CpGs in exon 7 of *LRPAP1*	Hypermethylation correspond to poor responsiveness to etanercept	[7]

correlated with erosive RA [51]. Second, a recent study by Rodriguez-Ubreva and de la Calle et al. identified a number of DNA methylation alterations in monocytes that directly correlated with RA disease activity score, namely DAS28 [10]. Furthermore, the DNA methylation levels of only three CpGs were able to stratify DAS28 with high accuracy [10]. Hence, these results provide a molecular alternative to the traditional method of evaluating RA disease activity. In the context of psoriasis, an inverse correlation between histone H4 acetylation and disease activity has been observed in isolated PBMCs, again highlighting the potential use of epigenetic alterations to predict disease activity [31]. Finally, several studies have linked aberrant epigenomes with differential response to conventional drug therapy. In one study, the authors identified 21 CpGs in isolated T cells, whose methylation levels were able to predict the response of newly diagnosed RA patients to DMARD with high sensitivity and specificity [52]. Correlation between response to therapy and DNA methylation was also observed by Plant and colleagues, in which hypermethylation of two CpGs located within exon 7 of the *LRPAP1* gene were reported to correlate with non-responsiveness of RA patients to etanercept [7].

Altogether, in the past few years, several prominent studies have provided new insights into the possible use of epigenetic alterations in immune cells as potential biomarkers in the context of autoimmune diseases, especially in RA. These new discoveries provide a step forward in the direction of a personalized approach in the diagnosis and treatment of patients.

Targeting epigenetic defects

Unlike genetic mutations, epigenetic alterations are potentially reversible processes which make them especially attractive as targets for drug therapy. In the context of several types of cancer, drug targeting of epigenetic modifications has been very successful in the past decade, and this has allowed the development of more specific and efficient therapy, loosely termed epi-drugs, specifically aimed to reverse aberrant epigenetic changes in cancer cells [53, 54]. More recently, several types of therapy targeting epigenetic modifications have been trialed in rheumatic diseases. This is of particular interest given the emerging role of epigenetic dysregulation in mediating aberrant phenotypes of immune cells in these diseases, shedding light to novel approaches in patient treatment and symptom management. Histone deacetylase inhibitors (HDACi) were one of the first

Table 2 Therapeutic treatments targeting epigenetic deregulation in rheumatic diseases.

Disease	Compound	Target	Molecular and clinical effects	Phase	Ref
sJIA	IT2357 (Givinostat)	Class I and II HDACi	Reduction in joint pain and inflammation	Approved	[55, 56]
SLE	Trichostatin A (TSA)	Class I and II HDACi	Downregulate aberrant expression of *IL-10*, *CD154*, and *IRF5*	Preclinical	[57–59]
SLE, RA	Vorinostat (SAHA)	HDAC1 and HDAC3 inhibitor	Inhibits NO expression in lupus-like mouse model; suppresses p38 and chemotaxis of RA synovial fibroblasts; induces apoptosis of RA FLS.	Preclinical	[60–62]
RA	MS-275	HDAC1 and HDAC3 inhibitor	Suppresses p38 signaling in synovial fibroblasts and induces apoptosis of FLS	Preclinical	[61, 62]
SLE	panobinostat	Pan HDACi	Significantly reduces auto-reactive plasma cell numbers	Preclinical	[63]
RA	MPT0G009	Pan HDACi	Anti-inflammatory effects and inhibits bone destruction	Preclinical	[64]

therapeutic agents to be extensively explored as potential treatments as the expression of several HDACs is increased in various rheumatic diseases [55, 65, 66] (see examples in Table 2). Givinostat, a class I and II HDAC inhibitor, has shown good efficacy in the treatment of systemic juvenile idiopathic arthritis (sJIA) during a phase II clinical trial, in which 17 patients displayed reduction in disease activity and joint inflammation compared to pretreatment, and has now been granted orphan drug designation in the European Union [56, 57]. Although no HDAC inhibitors have been tested in SLE and RA clinical trials, several have shown promising results in preclinical studies. One HDAC inhibitor, Trichostatin A (TSA), was first observed to modulate *IL-10* and *CD154* gene expression in SLE T cells [58]. Concordantly, treatment of TSA in a lupus-like mice model proved to be effective in reducing pathologic glomerular disease by increase T regulatory cells [59], and a more recent study by Shu et al. showed that TSA was able to downregulate aberrant expression of *IRF5* in PBMCs isolated from SLE patients [67]. Another HDAC inhibitor, varinostat, or SAHA, approved for the treatment of T-cell lymphoma and undergoing clinical trials for other malignancies, has been observed to modulate pathological inflammation in several experimental autoimmune disease models, including ANCA-associated vasculitis [68], autoimmune uveoretinitis [69], autoimmune encephalomyelitis [70], and type I diabetes [60]. The efficacy of SAHA has also been tested in the context of SLE and RA. A study by Reilly et al. showed that SAHA modulated the production of inflammatory mediators by inhibiting inducible NO expression in a lupus-like mouse model [61]. In the study of RA, SAHA has been observed to exert antirheumatic activities on RA synovial fibroblasts by suppressing the p38 MAPK signaling pathway, as well as inhibiting monocyte-mediated chemotaxis [71]. In accordance, a more recently study showed that SAHA was also effective in inducing apoptosis of RA-isolated fibroblast-like synoviocytes (FLS) by inhibiting NF-κB activation and increasing ROS production [62]. Finally, other HDAC inhibitors, including MS-275 [63, 71], panobinostat [64], and MPT0G009 [72] show promising results in a preclinical setting for the treatment of autoimmune diseases.

Collectively, these results demonstrate that the use of HDAC inhibitors may be a promising possibility for the treatment of autoimmune diseases. However, more studies are required to evaluate the safety and efficacy of these epi-drugs, and given the heterogeneous nature of these diseases, it is possible that the one-drug-fits-all model may not be applicable, and a more personalized approach is urgently required for autoimmune and autoinflammatory diseases.

References

[1] Schübeler D. Function and information content of DNA methylation. Nature 2015;517:321–6. https://doi.org/10.1038/nature14192.
[2] Kouzarides T. Chromatin modifications and their function. Cell 2007;128:693–705. https://doi.org/10.1016/j.cell.2007.02.005.

[3] Firestein GS, McInnes IB. Immunopathogenesis of rheumatoid arthritis. Immunity 2017;46:183–96. https://doi.org/10.1016/j.immuni.2017.02.006.

[4] Smolen JS, Aletaha D, Barton A, et al. Rheumatoid arthritis. Nat Rev Dis Primers 2018;4:18001. https://doi.org/10.1038/nrdp.2018.1.

[5] Liu Y, Aryee MJ, Padyukov L, et al. Epigenome-wide association data implicate DNA methylation as an intermediary of genetic risk in rheumatoid arthritis. Nat Biotechnol 2013;31:142–7. https://doi.org/10.1038/nbt.2487.

[6] van Steenbergen HW, Luijk R, Shoemaker R, et al. Differential methylation within the major histocompatibility complex region in rheumatoid arthritis: a replication study. Rheumatology (Oxford) 2014;53:2317–8. https://doi.org/10.1093/rheumatology/keu380.

[7] Plant D, Webster A, Nair N, et al. Differential methylation as a biomarker of response to etanercept in patients with rheumatoid arthritis. Arthritis Rheumatol 2016;68:1353–60. https://doi.org/10.1002/art.39590.

[8] de Andres MC, Perez-Pampin E, Calaza M, et al. Assessment of global DNA methylation in peripheral blood cell subpopulations of early rheumatoid arthritis before and after methotrexate. Arthritis Res Ther 2015;17:233. https://doi.org/10.1186/s13075-015-0748-5.

[9] Julià A, Absher D, López-Lasanta M, et al. Epigenome-wide association study of rheumatoid arthritis identifies differentially methylated loci in B cells. Hum Mol Genet 2017;26:2803–11. https://doi.org/10.1093/hmg/ddx177.

[10] Rodríguez-Ubreva J, de la Calle-Fabregat C, Li T, et al. Inflammatory cytokines shape a changing DNA methylome in monocytes mirroring disease activity in rheumatoid arthritis. Ann Rheum Dis Published Online First: 1 August 2019. https://doi.org/10.1136/annrheumdis-2019-215355.

[11] Gillespie J, Savic S, Wong C, et al. Histone deacetylases are dysregulated in rheumatoid arthritis and a novel HDAC3-selective inhibitor reduces IL-6 production by PBMC of RA patients. Arthritis Rheum Published Online First: 27 September 2011. https://doi.org/10.1002/art.33382.

[12] Li Y, Zhou M, Lv X, et al. Reduced activity of HDAC3 and increased acetylation of histones H3 in peripheral blood mononuclear cells of patients with rheumatoid arthritis. J Immunol Res 2018;2018:7313515. https://doi.org/10.1155/2018/7313515.

[13] Asadipour M, Hassan-Zadeh V, Aryaeian N, et al. Histone variants expression in peripheral blood mononuclear cells of patients with rheumatoid arthritis. Int J Rheum Dis 2018;21:1831–7. https://doi.org/10.1111/1756-185X.13126.

[14] Boehncke WH, Schön MP. Psoriasis: electrochemical behaviour of Tin (II) chloride as a solid state ionic conductor. Lancet 2015;4:169–77. https://doi.org/10.1016/S0140-6736(14)61909-7.

[15] Parisi R, Symmons DPM, Griffiths CEM, et al. Global epidemiology of psoriasis: a systematic review of incidence and prevalence. J Invest Dermatol 2013;133:377–85. https://doi.org/10.1038/jid.2012.339.

[16] Ritchlin CT, Colbert RA, Gladman DD. Psoriatic arthritis. N Engl J Med 2017;376:2095–6. https://doi.org/10.1056/NEJMc1704342.

[17] Grine L, Dejager L, Libert C, et al. An inflammatory triangle in psoriasis: TNF, type I IFNs and IL-17. Cytokine Growth Factor Rev 2015;26:25–33. https://doi.org/10.1016/j.cytogfr.2014.10.009.

[18] Christophers E, Metzler G, Röcken M. Bimodal immune activation in psoriasis. Br J Dermatol 2014;170:59–65. https://doi.org/10.1111/bjd.12631.

[19] Leonardi C, Matheson R, Zachariae C, et al. Anti-interleukin-17 monoclonal antibody ixekizumab in chronic plaque psoriasis. N Engl J Med 2012;366:1190–9. https://doi.org/10.1056/NEJMoa1109997.

[20] Chiricozzi A, Romanelli P, Volpe E, et al. Scanning the immunopathogenesis of psoriasis. Int J Mol Sci 2018;19. https://doi.org/10.3390/ijms19010179.

[21] Nestle FO, Di Meglio P, Qin JZ, et al. Skin immune sentinels in health and disease. Nat Rev Immunol 2009;9:679–91. https://doi.org/10.1038/nri2622.

[22] Mattozzi C, Salvi M, D'Epiro S, et al. Importance of regulatory T cells in the pathogenesis of psoriasis: review of the literature. Dermatology 2013;227:134–45. https://doi.org/10.1159/000353398.

[23] Yang L, Li B, Dang E, et al. Impaired function of regulatory T cells in patients with psoriasis is mediated by phosphorylation of STAT3. J Dermatol Sci 2016;81:85–92. https://doi.org/10.1016/j.jdermsci.2015.11.007.

[24] Soler DC, McCormick TS. The dark side of regulatory T cells in psoriasis. J Invest Dermatol 2011;131:1785–6. https://doi.org/10.1038/jid.2011.200.

[25] Zhang P, Su Y, Chen H, et al. Abnormal DNA methylation in skin lesions and PBMCs of patients with psoriasis vulgaris. J Dermatol Sci 2010;60:40–2. https://doi.org/10.1016/j.jdermsci.2010.07.011.

[26] Gervin K, Vigeland MD, Mattingsdal M, et al. DNA methylation and gene expression changes in monozygotic twins discordant for psoriasis: identification of epigenetically dysregulated genes. PLoS Genet 2012;8:https://doi.org/10.1371/journal.pgen.1002454.

[27] Han J, Park SG, Bae JB, et al. The characteristics of genome-wide DNA methylation in naïve CD4$^+$ T cells of patients with psoriasis or atopic dermatitis. Biochem Biophys Res Commun 2012;422:157–63. https://doi.org/10.1016/j.bbrc.2012.04.128.

[28] Park GT, Han J, Park SG, et al. DNA methylation analysis of CD4$^+$ T cells in patients with psoriasis. Arch Dermatol Res 2014;306:259–68. https://doi.org/10.1007/s00403-013-1432-8.

[29] Brandt D, Sergon M, Abraham S, et al. TCR$^+$ CD3$^+$ CD4$^-$ CD8$^-$ effector T cells in psoriasis. Clin Immunol 2017;181:51–9. https://doi.org/10.1016/j.clim.2017.06.002.

[30] Ovejero-Benito MC, Reolid A, Sánchez-Jiménez P, et al. Histone modifications associated with biological drug response in moderate-to-severe psoriasis. Exp Dermatol 2018;27:1361–71. https://doi.org/10.1111/exd.13790.

[31] Zhang P, Su Y, Zhao M, et al. Abnormal histone modifications in PBMCs from patients with psoriasis vulgaris. Eur J Dermatol 2011;21:552–7. https://doi.org/10.1684/ejd.2011.1383.

[32] Ekman A-K, Verma D, Fredrikson M, et al. Genetic susceptibility of NLRP1 in psoriasis. Br J Dermatol Published Online First: 2014. https://doi.org/10.1111/bjd.13178.

[33] Tsoi LC, Spain SL, Knight J, et al. Identification of 15 new psoriasis susceptibility loci highlights the role of innate immunity. Nat Genet 2012;44:1341–8. https://doi.org/10.1038/ng.2467.

[34] Wang Y, Edelmayer R, Wetter J, et al. Monocytes/Macrophages play a pathogenic role in IL-23 mediated psoriasis-like skin inflammation. Sci Rep 2019;9:5310. https://doi.org/10.1038/s41598-019-41655-7.

[35] Wang CQF, Suárez-Fariñas M, Nograles KE, et al. IL-17 induces inflammation-Associated gene products in blood monocytes, and treatment with ixekizumab reduces their expression in psoriasis patient blood. J Invest Dermatol 2014;134:2990–3. https://doi.org/10.1038/jid.2014.268.

[36] Aksentijevich I, Centola M, Deng Z, et al. Ancient missense mutations in a new member of the RoRet gene family are likely to cause familial Mediterranean fever. Cell 1997;90:797–807. https://doi.org/10.1016/S0092-8674(00)80539-5.

[37] Masumoto J, Taniguchi S, Ayukawa K, et al. ASC, a novel 22-kDa protein, aggregates during apoptosis of human promyelocytic leukemia HL-60 cells. J Biol Chem 1999;274:33835–8. https://doi.org/10.1074/jbc.274.48.33835.

[38] Shahbaznejad L, Raeeskarami S-R, Assari R, et al. Familial mediterranean gene (MEFV) mutation in parents of children with familial mediterranean fever: What are the exceptions? Int J Inflam 2018;2018:1–6. https://doi.org/10.1155/2018/1902791.

[39] Cekin N, Akyurek ME, Pinarbasi E, et al. MEFV mutations and their relation to major clinical symptoms of familial mediterranean fever. Gene 2017;626:9–13. https://doi.org/10.1016/j.gene.2017.05.013.

[40] Ben-Zvi I, Brandt B, Berkun Y, et al. The relative contribution of environmental and genetic factors to phenotypic variation in familial Mediterranean fever (FMF). Gene 2012;491:260–3. https://doi.org/10.1016/j.gene.2011.10.005.

[41] Aypar E, Ozen S, Okur H, et al. Th1 polarization in Familial Mediterranean fever. J Rheumatol 2003;30:2011–3 0315162X-30-2011.

[42] Kirectepe AK, Kasapcopur O, Arisoy N, et al. Analysis of MEFV exon methylation and expression patterns in familial Mediterranean fever. BMC Med Genet 2011;12:105. https://doi.org/10.1186/1471-2350-12-105.

[43] Erdem GC, Erdemir S, Abaci I, et al. Alternatively spliced MEFV transcript lacking exon 2 and its protein isoform pyrin-2d implies an epigenetic regulation of the gene in inflammatory cell culture models. Genet Mol Biol; 40: 688–97. https://doi.org/10.1590/1678-4685-GMB-2016-0234

[44] Kirectepe AK, Erdem GC, Senturk N, et al. Increased expression of exon 2 deleted MEFV transcript in familial Mediterranean fever patients. Int J Immunogenet 2011;38:327–9. https://doi.org/10.1111/j.1744-313X.2011.01015.x.

[45] Lamkanfi M, Dixit VM. Mechanisms and functions of inflammasomes. Cell 2014;157:1013–22. https://doi.org/10.1016/j.cell.2014.04.007.

[46] Takeuchi O, Akira S. Pattern recognition receptors and inflammation. Cell 2010;140:805–20 S0092-8674(10)00023-1 [pii]\r10.1016/j.cell.2010.01.022.

[47] Aksentijevich I, Putnam CD, Remmers EF, et al. The clinical continuum of cryopyrinopathies: novel CIAS1 mutations in North American patients and a new cryopyrin model. Arthritis Rheum 2007;56:1273–85. https://doi.org/10.1002/art.22491.

[48] Aubert P, Suárez-Fariñas M, Mitsui H, et al. Homeostatic tissue responses in skin biopsies from NOMID patients with constitutive overproduction of IL-1b. PLoS One Published Online First: 2012. https://doi.org/10.1371/journal.pone.0049408

[49] Vento-Tormo R, Álvarez-Errico D, Garcia-Gomez A, et al. DNA demethylation of inflammasome-associated genes is enhanced in patients with cryopyrin-associated periodic syndromes. J Allergy Clin Immunol 2017;139. https://doi.org/10.1016/j.jaci.2016.05.016. 202–211.e6.

[50] Chen S, Pu W, Guo S, et al. Genome-wide DNA methylation profiles reveal common epigenetic patterns of interferon-related genes in multiple autoimmune diseases. Front Genet 2019;10:223. https://doi.org/10.3389/fgene.2019.00223.

[51] Mok A, Rhead B, Holingue C, et al. Hypomethylation of CYP2E1 and DUSP22 promoters associated with disease activity and erosive disease among rheumatoid arthritis patients. Arthritis Rheumatol (Hoboken, NJ) 2018;70:528–36. https://doi.org/10.1002/art.40408.

[52] Glossop JR, Nixon NB, Emes RD, et al. DNA methylation at diagnosis is associated with response to disease-modifying drugs in early rheumatoid arthritis. Epigenomics 2017;9:419–28. https://doi.org/10.2217/epi-2016-0042.

[53] Pfister SX, Ashworth A. Marked for death: targeting epigenetic changes in cancer. Nat Rev Drug Discov 2017;16:241–63. https://doi.org/10.1038/nrd.2016.256.

[54] Park JW, Han J-W. Targeting epigenetics for cancer therapy. Arch Pharm Res 2019;42:159–70. https://doi.org/10.1007/s12272-019-01126-z.

[55] Toussirot E, Abbas W, Khan KA, et al. Imbalance between HAT and HDAC activities in the PBMCs of patients with ankylosing spondylitis or rheumatoid arthritis and influence of HDAC inhibitors on TNF alpha production. PLoS One 2013;8:e70939. https://doi.org/10.1371/journal.pone.0070939.

[56] Vojinovic J, Damjanov N. HDAC inhibition in rheumatoid arthritis and juvenile idiopathic arthritis. Mol Med 2011;17:397–403. https://doi.org/10.2119/molmed.2011.00030.

[57] Vojinovic J, Damjanov N, D'Urzo C, et al. Safety and efficacy of an oral histone deacetylase inhibitor in systemic-onset juvenile idiopathic arthritis. Arthritis Rheum 2011;63:1452–8. https://doi.org/10.1002/art.30238.

[58] Mishra N, Brown DR, Olorenshaw IM, et al. Trichostatin A reverses skewed expression of CD154, interleukin-10, and interferon-gene and protein expression in lupus T cells. Proc Natl Acad Sci 2001;98:2628–33. https://doi.org/10.1073/pnas.051507098.

[59] Reilly CM, Thomas M, Gogal R, et al. The histone deacetylase inhibitor trichostatin A upregulates regulatory T cells and modulates autoimmunity in NZB/W F1 mice. J Autoimmun 2008;31:123–30. https://doi.org/10.1016/j.jaut.2008.04.020.

[60] Cabrera SM, Colvin SC, Tersey SA, et al. Effects of combination therapy with dipeptidyl peptidase-IV and histone deacetylase inhibitors in the non-obese diabetic mouse model of type 1 diabetes. Clin Exp Immunol 2013;172:375–82. https://doi.org/10.1111/cei.12068.

[61] Reilly CM, Mishra N, Miller JM, et al. Modulation of renal disease in MRL/*lpr* mice by suberoylanilide hydroxamic acid. J Immunol 2004;173:4171–8. https://doi.org/10.4049/jimmunol.173.6.4171.

[62] Chen H, Pan J, Wang J, et al. Suberoylanilide hydroxamic acid, an inhibitor of histone deacetylase, induces apoptosis in rheumatoid arthritis fibroblast-like synoviocytes. Inflammation 2016;39:39–46. https://doi.org/10.1007/s10753-015-0220-3.

[63] Choo Q-Y, Ho PC, Tanaka Y, et al. Histone deacetylase inhibitors MS-275 and SAHA induced growth arrest and suppressed lipopolysaccharide-stimulated NF-B p65 nuclear accumulation in human rheumatoid arthritis synovial fibroblastic E11 cells. Rheumatology 2010;49:1447–60. https://doi.org/10.1093/rheumatology/keq108.

[64] Waibel M, Christiansen AJ, Hibbs ML, et al. Manipulation of B-cell responses with histone deacetylase inhibitors. Nat Commun 2015;6:6838. https://doi.org/10.1038/ncomms7838.

[65] Kenneth LW, Edward T, Kayihura M, et al. Modulation of histone deacetylases (HDACs) expression in patients with and without systemic lupus erythematosus: possible drug targets for treatment. J Rheum Dis Treat 2018;4. https://doi.org/10.23937/2469-5726/1510060.

[66] Angiolilli C, Kabala PA, Grabiec AM, et al. Histone deacetylase 3 regulates the inflammatory gene expression programme of rheumatoid arthritis fibroblast-like synoviocytes. Ann Rheum Dis 2017;76:277–85. https://doi.org/10.1136/annrheumdis-2015-209064.

[67] Shu J, Li L, Zhou L-B, et al. IRF5 is elevated in childhood-onset SLE and regulated by histone acetyltransferase and histone deacetylase inhibitors. Oncotarget Published Online First: May 2017. doi:10.18632/oncotarget.17586

[68] Dooley D, van Timmeren MM, O'Reilly VP, et al. Alkylating histone deacetylase inhibitors may have therapeutic value in experimental myeloperoxidase-ANCA vasculitis. Kidney Int 2018;94:926–36. https://doi.org/10.1016/j.kint.2018.05.028.

[69] Fang S, Meng X, Zhang Z, et al. Vorinostat modulates the imbalance of T cell subsets, suppresses macrophage activity, and ameliorates experimental autoimmune uveoretinitis. Neuromolecular Med 2016;18:134–45. https://doi.org/10.1007/s12017-016-8383-0.

[70] Ge Z, Da Y, Xue Z, et al. Vorinostat, a histone deacetylase inhibitor, suppresses dendritic cell function and ameliorates experimental autoimmune encephalomyelitis. Exp Neurol 2013;241:56–66. https://doi.org/10.1016/j.expneurol.2012.12.006.

[71] Choo Q-Y, Ho P, Tanaka Y, et al. The histone deacetylase inhibitors MS-275 and SAHA suppress the p38 mitogen-activated protein kinase signaling pathway and chemotaxis in rheumatoid arthritic synovial fibroblastic E11 cells. Molecules 2013;18:14085–95. https://doi.org/10.3390/molecules181114085.

[72] Hsieh I-N, Liou J-P, Lee H-Y, et al. Preclinical anti-arthritic study and pharmacokinetic properties of a potent histone deacetylase inhibitor MPT0G009. Cell Death Dis 2014;5:e1166. https://doi.org/10.1038/cddis.2014.133.

This page intentionally left blank

CHAPTER 12

Epigenetics of allergies: From birth to childhood

Avery DeVries[a], **Donata Vercelli**[a,b,c]
[a]Asthma & Airway Disease Research Center, University of Arizona, Tucson, AZ, United States
[b]Department of Cellular and Molecular Medicine, University of Arizona, Tucson, AZ, United States
[c]Arizona Center for the Biology of Complex Diseases, University of Arizona, Tucson, AZ, United States

Contents

Neonatal DNA methylation profiles as predictors of the trajectory to asthma and allergy 273
DNA methylation profiles in patients with childhood asthma and allergy 276
    DNA methylation in the airways 278
What have we learned so far? 279
References 281

Allergic diseases are among the most common chronic diseases that affect both children and adults. They are typically characterized by skewed type-2 immune responses, with increased expression of IL-4, IL-5, and IL-13 by T cells and type-2 innate lymphoid cells and vigorous production of IgE to otherwise innocuous environmental antigens (allergens). In asthma, a disease often associated with allergic manifestations, these immune alterations are thought to reflect primarily abnormal responses of the airway epithelium to environmental stimuli, leading to excessive release of IL-33, IL-25, and TSLP, innate cytokines that activate type-2 responses in the respiratory mucosa [1–3]. Importantly, asthma—a disease characterized by recurrent, reversible bronchial obstruction, and chronic inflammation of the airways—is phenotypically heterogeneous and includes multiple, distinct endotypes. Thus, even though type-2 inflammatory responses are typically critical for asthma pathogenesis, additional cellular and molecular mechanisms (e.g., increased IL-17 production and neutrophilia [4], reduced T regulatory activity [2, 5]) underpin the vast spectrum of asthma phenotypes.

Many efforts have been made to understand the genetic basis of allergic disease susceptibility because of their strong family history and elusive pathogenesis. Both candidate gene and genome-wide surveys were performed. In asthma, genome-wide association studies (GWAS) successfully identified a number of plausible risk-associated genes [for instance, *IL33*, *IL1RL1* (a subunit of the IL33 receptor), *TSLP*, and type-2 cytokines], and a number of novel candidates—the 17q locus first and foremost [6]. Variants

Epigenetics of the Immune System
https://doi.org/10.1016/B978-0-12-817964-2.00012-5

© 2020 Elsevier Inc.
All rights reserved.

specifically associated with childhood-onset, as opposed to adult-onset, asthma were identified [7], and unexpected similarities between seemingly unrelated diseases (e.g., childhood asthma and Crohn's disease) [8] also emerged. Moreover, diseases with opposing immune mechanisms and mutually exclusive phenotypes (e.g., atopic dermatitis and psoriasis) were found to be associated with opposite risk alleles within shared susceptibility loci [9]. Despite these successes, most genome-wide associations have only modest effect sizes, and only a small proportion of phenotypic variability can be explained by all GWAS associations combined [10].

The "missing heritability" found for most complex diseases has provided a rationale to search for other sources of phenotypic variance, particularly among factors related to the environment and development. Birth cohorts in particular are designed to explore the time in early life that represents a critical developmental window for allergic disease risk. Perinatal and postnatal factors (such as infections, mode of delivery, preterm birth, birthweight and exposures to microbes, air pollution, and tobacco smoke [11–13]) have been shown to modify a child's risk for allergy. For example, early life exposure to microbe-rich farming environments has been shown to protect against asthma and allergies [14], while early life wheezing illnesses due to rhinovirus or respiratory syncytial virus [11, 12] are associated with increased risk. These findings emphasize that exposures occurring in early life and even in utero can have a lasting impact on a child's trajectory to asthma and allergy.

Because allergic disease susceptibility and pathogenesis involve genetic, environmental, and developmental factors, much research is focusing on epigenetic mechanisms in an effort to gain insights into the inception and course of these diseases. Indeed, epigenetic mechanisms regulate gene expression in a manner independent of changes to the DNA sequence, are involved in the timed unfolding of developmental programs, and provide the plasticity required to respond to environmental stimuli. Fundamental epigenetic mechanisms include DNA methylation and posttranslational modifications of histone tails. DNA methylation is most studied in human populations because it is a robust phenotype that reflects gene regulatory events and can be quantitatively assessed using widely available commercial platforms [15, 16]. Most studies so far have characterized DNA methylation *at the time of disease*, a trait that is readily measurable but also inherently noisy. Moreover, and most importantly, it is virtually impossible to determine whether a given disease-associated pattern of DNA methylation is a cause or a consequence of the disease [17]. This is why a growing number of studies are now exploring DNA methylation profiles *at birth* and their contribution to subsequent *disease trajectories*. Studies of this kind are the primary, albeit not the only, focus of this review. The designs of all the studies discussed in this chapter are shown in Tables 1 and 2.

Table 1 Epigenome-wide DNA methylation studies of asthma and allergy focused on birth.

Primary outcome(s)	Discovery cohort(s)	Sample size (*n*)	Secondary analyses/ replication	Tissue	Method (platform)	Ref No.
Asthma (2–9 years)	IIS	36 (genome-wide discovery phase) 60 (targeted analysis phase)	Replication of SMAD3 DMR in CBMCs from two independent populations (MAAS, $n=30$; and COAST, $n=28$)	CBMC	NimbleGen 2.1M Human Promoter Deluxe	[18]
Childhood lung function: FEV_1, FEV_1/FVC, and FEF_{75} (7–13 years)	ALSPAC, Generation R, INMA, CHS, Project Viva	1688	**Secondary analyses in cord blood**: childhood asthma (Generation R, $n=710$) and adolescent lung function (ALSPAC, $n=542$) **Secondary analyses in peripheral blood**: adult lung function (Rotterdam Study, $n=1191$), and COPD (Rotterdam Study, $n=1309$)	Cord blood	450K	[19]
Asthma (5–7 years)	ALSPAC, CHS, EDEN, Generation R, GOYA, MoBa1, MoBa2, NEST	3572	N/A	Cord blood/ CBMC	450K	[20]
Total serum IgE Atopic sensitization (food and environmental), Asthma (mean age=7.8 years)	Project Viva	485	N/A	Cord blood	450K	[21]

450K, Illumina Human Methylation450 BeadChip; *ALSPAC*, Avon Longitudinal Study of Parents and Children; *CBMC*, cord blood mononuclear cells; *CHS*, Children's Health Study; *COAST*, Childhood Origins of Asthma Study; *EDEN*, Etude des Determinants pre et post natals du developpement et de la sante de l'Enfant; *GOYA*, Genetics of Overweight Young Adults; *IIS*, Infant Immune Study; *INMA*, Infancia y Medio Ambiente; *MAAS*, Manchester Asthma and Allergy Study; *MoBa*, Norwegian Mother and Child; *NEST*, Newborn Epigenetics Study.

Table 2 Epigenome-wide DNA methylation studies of asthma and allergy in childhood.

Primary outcome(s)	Discovery cohort(s)	Sample size (*n*)	Secondary analyses/ replication	Age	Tissue	Method (platform)	Ref No.
Asthma (7–17 years)	BAMSE EpiGene, BAMSE MeDALL, CHOP, GALA II, ICAC, NFBC 1986, PIAMA, Raine study, STOPPA	2862	Replication in purified eosinophils from SLSJ ($n=24$, 2–56 years) and nasal epithelium from PIAMA ($n=455$, 16 years) and ICAC ($n=72$, 10–12 years)	7–17 years	Whole blood/ PBMC	450K	[1]
Total serum IgE Atopic sensitization, Environmental and food allergen sensitization, Asthma (mean age = 7.8 years)	Project Viva	120 (early childhood) 408 (mid-childhood)	N/A Replication in peripheral blood from GACRS ($n = 159$, $6 - 14$ years)	Early childhood (mean age = 3.4 years) Mid-childhood (mean age = 7.8 years)	Whole blood	450K	[2]
Asthma (4–5 or 8 years)	BAMSE, EDEN, INMA, PIAMA BIB, ECA, INMA, RHEA, ROBBIC, PIAMA	817 (4 – 5 years) 731 (8 years) 3196 (meta-analysis)	Replication in whole blood ($n=167$, 1–79 years) and purified eosinophils from SLSJ ($n=24$, 2–56 years) and nasal epithelium from PIAMA ($n=455$, 16 years)	4–5 or 8 years	Whole blood	450K	[3]
Allergic sensitization (10 and 18 years)	IoW	376 (DNAm at both ages)	Replication in whole blood from BAMSE ($n=267$, 16 years)	10 and 18 years	Whole blood	450K & EPIC	[4]
Allergic asthma (10–12 years)	ICAC	72	Replication in nasal epithelium (National Jewish Health, $n=24$, 24–74 years)	10–12 years	Nasal epithelium	450K	[5]
Atopic sensitization or atopic asthma (9–20 years)	EVA-PR	483	Replication in nasal epithelium from ICAC ($n=72$, 10–12 years) and PIAMA ($n=432$, 16 years)	9–20 years	Nasal epithelium	450K	[6]

450K, Illumina Human Methylation450 BeadChip; *BAMSE*, Barn/Children, Allergy, Milieu, Stockholm, Epidemiology; *BIB*, Born in Bradford; *ECA*, Environment and Childhood Asthma study in Oslo; *EDEN*, Etude des Determinants pre et post natals du developpement et de la sante de l'Enfant; *EPIC*, Illumina MethylationEPIC BeadChip; *EVA-PR*, Epigenetic Variation and Childhood Asthma in Puerto Ricans; *GALA II*, Genes-environments & Admixture in Latino Americans; *ICAC*, Inner-City Asthma Consortium; *INMA*, Infancia y Medio Ambiente; *IoW*, Isle of Wight; *MeDALL*, Mechanisms of the Development of ALLergy; *PBMC*, peripheral blood mononuclear cells; *PIAMA*, Prevention and Incidence of Asthma and Mite Allergy; *RHEA*, Mother-Child Cohort in Crete; *ROBBIC*, Rome and Bologna Birth Cohorts; *SLSJ*, Saguenay-Lac-Saint-Jean.

Neonatal DNA methylation profiles as predictors of the trajectory to asthma and allergy

The approach we took to identify epigenetic determinants of childhood asthma is driven primarily by biological consideration and thus, interprets DNA methylation differences as telltale markers of functionally relevant pathways. Our recent epigenome-wide analysis in cord blood mononuclear cells identified 589 differentially methylated regions (DMRs) associated with asthma at age 2–9 years in children enrolled in the Infant Immune Study (IIS), a longitudinal birth cohort closely monitored for asthma within the first decade of life [18]. When a network of DMR-containing genes was constructed to decipher the functional relationships among the DMRs and the phenotype, we found that a DMR mapping to the *SMAD3* promoter was the most connected node in the network, and the proinflammatory cytokine IL-1β was a prominent upstream regulator of network genes. Because *SMAD3* is a replicated asthma gene in GWAS [22–24] and acts as the master regulator of TGF-β signaling, we examined the *SMAD3* DMR at higher resolution. Bisulfite sequencing of this region in 60 IIS neonates showed that *SMAD3* promoter methylation at birth was significantly and selectively increased in children who became asthmatic and were born to asthmatic mothers. This finding was replicated in the Manchester Allergy and Asthma Study and the Childhood Origins of ASThma Study. A meta-analysis reiterated the association between *SMAD3* promoter methylation and childhood asthma risk: for each 10% increase in *SMAD3* methylation, there was nearly a twofold increased risk of childhood asthma (meta-analysis odds ratio [OR] = 1.95, [95%CI: 1.23, 3.10], $P = 0.005$). In this high-resolution data set, we also sought a functional relationship between *SMAD3*, the primary hub in our asthma-associated gene network, and IL-1β, a proinflammatory cytokine involved in human asthma [25–27] and the strongest upstream regulator of the asthma-associated network. Previous studies in mouse models [28] support this connection. A positive relationship between *SMAD3* DMR methylation and neonatal IL-1β producing capacity was detected in IIS children born to asthmatic mothers, but not in children of non-asthmatic mothers.

Our study was the first to show that the trajectory to childhood asthma likely begins at birth (if not sooner) and may involve epigenetic modifications clustering in immunoregulatory (*SMAD3*) and proinflammatory (*IL1B*) pathways. Importantly, this trajectory appears to be influenced by the asthma status of the child's mother. By regulating TGF-β signaling, SMAD3 controls the balance between the asthma-promoting Th17 and the asthma-protective Treg programs, which is thought to be altered in childhood asthma [2, 5]. Moreover, TGF-β is a critical regulator of lung development [29]. Neonates who became asthmatic in the first decade of life had high *SMAD3* promoter methylation (an epigenetic configuration consistent with low *SMAD3* expression) and high IL-1β production. This combination may destabilize the Treg program, promote Th17 differentiation and inflammation [30], and ultimately favor the development of asthma.

It is unknown whether these mechanisms operate prenatally and/or perinatally, but the selective relationship between *SMAD3* methylation and IL-1β production among children of asthmatic mothers suggests that the in utero environment critically influences the epigenetic trajectory to childhood asthma.

Subsequent studies have sought to characterize the relations between immune epigenetic profiles at birth and asthma-related phenotypes. den Dekker et al. conducted an epigenome-wide meta-analysis of DNA methylation in 1688 cord blood cell samples from five cohorts in an attempt to better understand the epigenetic mechanisms associated with lung function during childhood [19]. Initial analyses identified 59 DMRs [defined as regions containing at least 2 CpGs within a 500-bp window and having a false discovery rate (FDR)-adjusted P-value < 0.05] associated with lung function parameters (FEV_1, FEV_1/FVC, or FEF_{75}) at age 7–13 years. Secondary analyses within individual cohorts then assessed whether these DMRs were also associated with other respiratory phenotypes across the subjects' life span. Of the original 59 DMRs, 18 were associated with asthma during childhood (mean age 6 years, GenerationR study), 11 and 9 were associated with lung function during adolescence [mean age 15 years, Avon Longitudinal Study of Parents and Children (ALSPAC) study] and adulthood (mean age 66 years, Rotterdam study), and 9 were associated with COPD (mean age 66 years, Rotterdam study). Many of these DMRs were also associated with gene expression in blood samples [32 DMRs at age 4 (Infancia y Medio Ambiente, INMA) and 18 DMRs in adulthood (Rotterdam study)], supporting their putative regulatory function. In all, 43 DMRs were annotated to novel candidate genes. Among the 10 most significant DMRs, *HOXA5*, *CLCA1*, *TCL1A,* and *NUDT12* had been previously associated with lung development, respiratory diseases, and cellular immunity. An association between DNA methylation levels and adult COPD had been reported for four genes (*CBFA2T3*, *PADI4*, *LST1*, and *KCNQ1*). The authors considered their findings a replication of previous studies because 19 DMRs from the Inner-City Asthma Consortium (ICAC) cohort [31] were located within 500 kb of their DMRs. However, the ability of DNA methylation to regulate genes over such a distance remains unclear and may depend on the location of the DMR. Overall, these findings are consistent with the notion that changes in DNA methylation in early life (neonatal or even prenatal) may have long-lasting effects on respiratory disease. By focusing on differential methylation associated with lung function, these authors may have identified a core set of epigenetic mechanisms that remain important throughout life.

In another recent study, Reese and colleagues searched for asthma-associated differential methylation by performing a prospective analysis in newborns from eight cohorts and a cross-sectional analysis in 7- to 17-year-old children from nine cohorts [20]. In 3572 newborns, the authors identified 9 asthma-associated differentially methylated CpGs (DMCs) and 35 asthma-associated DMRs, defined as containing at least 2 CpGs within a 1000-bp window with FDR-adjusted P-values < 0.01. Methylation levels at the majority of these loci were also correlated with gene expression in at least one cohort. Among the asthma-associated DMCs identified in children, 20 and 128 were replicated in nasal epithelium from the Prevention and Incidence of Asthma and Mite Allergy (PIAMA) and ICAC cohorts, respectively, and 148 were replicated in purified eosinophils from the Saguenay-Lac-Saint-Jean (SLSJ) cohort.

While each DMC was significantly associated with asthma risk, none of the 35 neonatal DMRs contained significant DMCs, suggesting that these regions carried only weak associations. In contrast, the analyses performed in older children identified 179 asthma-associated DMCs and 36 DMRs, 21 of which contained at least one DMC. No DMC identified in newborns was significantly associated with asthma in older children, and only 6/179 DMCs identified in older children were nominally significant ($P < 0.05$) in newborns. However, pathways and biological functions enriched for asthma-associated differential methylation significantly overlapped between newborns and children. Newborn DMCs/DMRs were enriched for canonical pathways generally relevant to asthma, such as eNOS signaling, the inflammasome, and NFκB signaling ($P = 0.039$, $P = 0.037$, and $P = 0.046$, respectively). More pathways and biological functions were enriched in DMCs/DMRs from older children, including mTOR and ERK5 signaling ($P = 0.0012$ and $P = 0.0026$, respectively), which have been associated with asthma and pulmonary fibrosis. Interestingly, NF-κB activation and eNOS signaling pathways were also enriched in the DMCs/DMRs from older children ($P = 0.041$ and $P = 0.047$, respectively), suggesting a potential functional convergence of these findings. It is unclear why replication patterns differed at different ages in this study, but the fact that study samples were obtained from distinct cohorts presumably exposed to distinct environments both prenatally and postnatally may account for at least some of the reported inconsistencies.

On the theme of age, an interesting approach was recently leveraged to explore the connection between allergies and asthma and age-related epigenetic mechanisms. Increasing evidence suggests that "epigenetic clock" metrics based on DNA methylation are associated with complex diseases. Both DNA methylation age (DNAmAge, i.e., age estimated from DNA methylation profiles) and epigenetic age acceleration (i.e., a higher DNA methylation age relative to chronological age) have been linked cross sectionally with traits like physical and cognitive fitness [32] and have successfully predicted the subsequent onset of cancer and mortality [33–37]. Recently, this work has been expanded to include early life events. Studies have shown that prenatal and perinatal factors and exposures (e.g., maternal smoking during pregnancy and mode of delivery) are associated with higher DNAmAge in children [38], and epigenetic age acceleration is prospectively associated with complex traits (e.g., weight gain [39] and earlier puberty in girls [40]) during childhood and adolescence. To investigate the link between epigenetic clocks and asthma and allergies, published algorithms [41, 42] were used to estimate DNAmAge and epigenetic age acceleration rates at birth, early childhood, and mid-childhood within the Project Viva cohort. The association of these metrics with allergic phenotypes (asthma, total serum IgE, atopic sensitization, and environmental and food allergen sensitization) was subsequently tested. Mid-childhood DNAmAge was found to be associated cross sectionally with allergy and asthma phenotypes (1.2 higher odds of disease). Epigenetic age acceleration was also associated with disease (1.2 higher odds of disease for every 1-year increase). Interestingly, gender, ethnicity, and maternal smoking status did not modify the association between DNAmAge and disease, but maternal socioeconomic factors such as level of maternal education and marital status during pregnancy were significantly associated with DNAmAge in mid-childhood. Neither DNAmAge nor epigenetic age acceleration at birth or in early-childhood were significantly associated with allergic phenotypes in mid-childhood. The authors also estimated DNA methylation gestational age, which was also unrelated to allergic phenotypes in mid-childhood. Associations with atopic sensitization and asthma identified in mid-childhood were replicated in an independent cohort, and associations with IgE showed comparable estimates and direction of effects, even though they did not reach statistical significance.

These data were the first to suggest that DNAmAge and epigenetic age acceleration during mid-childhood are associated with allergic disease. While pathway-related analyses were not performed, the 353 methylation sites originally used to estimate the epigenetic clock were significantly enriched for pathways critical to immune cell function and organismal and tissue development [41]. The fact that a higher DNAmAge was associated with maternal socioeconomic factors during pregnancy points to a potentially unique trajectory to allergic disease.

DNA methylation profiles in patients with childhood asthma and allergy

Although the relationship between DNA methylation at birth and allergic diseases later in life is the primary focus of this review, here we will briefly turn our attention to studies of DNA methylation *concurrent with disease,* a topic that has been recently addressed by several high-profile papers. A comprehensive review of the field is also provided in Ref. [43].

A recent multiphase epigenome-wide study in multiple cohorts sought to understand how asthma-associated epigenetic profiles might be acquired overtime. In all, 35 asthma-associated DMCs were identified in the discovery phase [11 at age 4 years ($\text{FDR} < 0.10$) and an additional 24 at age 8 years (at a genome-wide significance threshold of $P < 1.14 \times 10^{-7}$)] [44]. In all, 14 of these sites were validated in a meta-analysis of 3196 individuals from 6 cohorts participating in the Mechanisms of the Development of ALLergy project and were selected for further analysis. All 14 DMCs were independently replicated in whole blood DNA from the SLSJ cohort ($n = 167$, age 1–79 years), but not in cord blood ($n = 1316$). The authors interpret these results as a reflection of postnatal environmental events and/or pathophysiological disease processes. However, the relationships between DNA methylation and environmental exposures such as early life viral infections or air pollution were not investigated, nor were these relationships examined in paired samples collected from individual subjects at birth and age 4 years.

Because adjusting DNA methylation levels for number of eosinophils attenuated the associations found in the discovery phase, the authors also assessed whether the CpGs in question were differentially methylated in isolated eosinophils. Using data from a previously published study [45], they confirmed that 11/14 asthma-associated DMCs were hypomethylated in eosinophils compared to other blood cell types (Fisher's exact test, $P = 1.92 \times 10^{-24}$). Moreover, data generated from isolated eosinophils in the SLSJ cohort showed that all 14 DMCs were hypomethylated in asthmatics compared to nonasthmatics. Interestingly, 5 of these 14 DMCs were also replicated in nasal epithelial cells from children in the PIAMA study, suggesting that the relevant genes may contribute to both immune and respiratory mechanisms of disease.

Further analyses demonstrated that DNA methylation levels at all 14 DMCs were associated with gene expression in whole blood, and hierarchical clustering identified three distinct clusters of expression patterns. The first cluster included genes expressed in eosinophils (*SIGLEC8*, *IL5RA*, and *PTGDR2*) and was inversely correlated with methylation levels at seven asthma-associated DMCs. The second cluster included *CCR7*, *CHMP7*, and *LEF1* and was strongly associated with seven DMCs. This cluster represented a naïve CD4 and CD8 T cell gene signature. The third cluster represented an effector/memory CD8 T cell and natural killer cell signature and included genes such as *TBX21*, *EOMES*, *CCL5*, and *GZMB*, which were negatively correlated with four DMCs. These results, along with the lack of replication in cord blood, suggest that certain epigenetic profiles are acquired overtime and raise the possibility that early life environmental exposures may elicit lasting epigenetic responses related to the asthma trajectory.

Few of the studies discussed so far have monitored the longitudinal relationship between DNA methylation and allergic disease within the *same* individuals or populations. A recent publication focused on 376 subjects from the Isle of Wight cohort to understand how preadolescence and postadolescence peripheral blood DNA methylation influences allergic sensitization and poly-sensitization [46]. Among 35 DMCs associated with allergic sensitization, 34 were hypomethylated in allergic subjects. Two DMCs within *ACOT7* showed the largest reduction in methylation between allergic and nonallergic individuals. Genes containing differential methylation associated with allergic sensitization were significantly enriched for biological processes involved in regulation of IL-5 production. Notably, 27/35 DMCs were replicated in an independent cohort [Barn/Children, Allergy, Milieu, Stockholm, Epidemiology (BAMSE), $n = 267$] at a significance of $P < 0.05$. Data from the BAMSE replication cohort were also used to assess whether methylation at the 35 DMCs was associated with gene expression. The strongest association was observed between cg10159529 in the *IL5RA* promoter and *IL5RA* expression ($P = 6.76 \times 10^{-20}$).

Because the authors focused on adolescence, they investigated whether allergy-associated differential methylation might be age and gender specific. Indeed, a proportion of allergic sensitization-associated DMCs had significant interactions with age, but none showed a significant interaction with gender. Moreover, the three CpGs whose methylation was associated with poly-sensitization had no significant interactions with age or gender. Together, these findings show that allergic sensitization is associated with DNA methylation during adolescence, and some of these associations are age specific. These findings also highlight how important it is to ascertain DNA methylation longitudinally in the same individuals, and suggest that different relationships may exist depending on the disease endotype.

DNA methylation in the airways

While for several years DNA methylation studies in allergy have relied on circulating immune cells, the most recent efforts have focused on DNA methylation in the airways. Investigators from the ICAC cohort were among the first to study DNA methylation in nasal epithelial cells as a surrogate for smaller airways epithelium. The authors identified a total of 186 allergic asthma-associated differentially methylated genes (119 DMRs representing 118 unique genes, and 118 single CpGs representing 107 unique genes) in nasal epithelial cells from 72 individuals, the majority of whom were African-Americans [47]. Among the significant hits, several genes (*ALOX15*, *CAPN14*, *HNMT*, and *POSTN*) had been previously associated with asthma and allergies. Notably, this study was one of the first to emphasize that peripheral blood cells may better model immune mechanisms of allergic disease, while nasal epithelial cells may better represent airway-based events. Indeed, only three DMRs identified in peripheral blood mononuclear cells from the same population [31] were replicated in nasal epithelium. In contrast, 67 of the 186 asthma-associated differentially methylated genes were replicated in nasal epithelial cells from adult Caucasian atopic asthmatics compared with non-asthmatic/nonatopic subjects.

Interestingly, 47 differentially methylated genes were inversely related to the expression of asthma-associated differentially expressed genes. An upstream regulator analysis on these 47 asthma-associated genes showed an enrichment for genes regulated by known asthma mediators, such as IL-4, IL-6, IL-13, IFN-γ, TGF-β, and the transcriptional regulator CIITA. Molecular interaction networks identified RIPK2 as the main hub, suggesting that both adaptive and innate immune pathways are important for allergic airway inflammation. Together, these findings suggest that genes whose expression is altered in asthmatic airways may be regulated by epigenetic mechanisms.

A more recent epigenome-wide analysis in nasal epithelial cells examined relationships between DNA methylation and atopy or atopic asthma in a population of Puerto Ricans ($n = 483$, age 9–20 years) [48]. A total of 8664 atopy-associated DMCs were identified at an FDR-adjusted $P\text{-value} < 0.01$. Interestingly, 5101 of these DMCs were also differentially methylated between individuals with and without atopic asthma and atopy. Among the top 30 atopy-associated DMCs, several mapped to or were near biologically plausible genes for atopy, including genes involved in epithelial barrier function (e.g., *CDHR3* and *CDH26*) and genes involved in epithelial integrity and immune regulation (e.g., *FBXL7*, *NTRK1*, and *SLC9A3*). The authors showed that 28 and 29 of the top 30 DMCs replicated ($P<0.01$) with the same direction of effect in the ICAC and PIAMA cohorts, respectively.

There was substantial overlap between epigenome-wide differential methylation and transcriptome-wide expression. In fact, 13 of the top 30 atopy-associated differentially methylated genes (including *SLC9A3*, *PCSK6*, *CDH26*, *FBXL7,* and *NTRK1*) were also differentially expressed in atopics and nonatopics. Moreover, 1570 DMC-gene expression pairs (within 10 kb of each other) emerged from an expression quantitative trait DNA methylation (eQTM) analysis, including 11 from the top 30 atopy-associated DMCs. The authors performed pathway analysis on a core set of genes ($n=724$, FDR-adjusted $P<0.01$ in the epigenome-wide, transcriptome-wide, and eQTM analyses) and identified an enrichment for immune regulation (Th1 and Th2 pathways, antigen presentation, regulatory T cell, and NK cell signaling pathways), gap junction signaling and epidermal growth factor receptor pathways.

Finally, a 30 CpG-panel was created and evaluated for its performance as an epigenetic classifier. Classification models performed well for atopy in the discovery population [area under the curve (AUC) 0.93–0.94, accuracy 85%–88%] and in both replication cohorts (AUC 0.74–0.92, accuracy 68%–82%). The classification models performed well also for atopic asthma in both the discovery cohort (AUC 0.95–1.00, accuracy 88%) and the replication cohorts (AUC 0.82–0.88, accuracy 86%), suggesting that these CpGs may broadly classify allergic disease.

The genes associated with differential methylation that are discussed in this review are shown in Fig. 1.

What have we learned so far?

The current landscape in asthma and allergy epigenetics is somewhat puzzling. Most of the studies discussed above relied on population samples of reasonable size and measured DNA methylation on the same platform, but overall very few differentially methylated loci were identified in each study, and even fewer consistently replicated across populations [43]. The reasons behind this state of affairs are probably several and complex. One possibility is that hidden heterogeneity may have masked or attenuated relationships that were significant (biologically as well as statistically) only in a subset of the study population. The interaction we found between *SMAD3* methylation, childhood and maternal asthma is a case in point [18]. Multi-cohort studies that maximize sample size are currently extremely popular, but the meta-analyses required to extract results from these studies seek commonalities rather

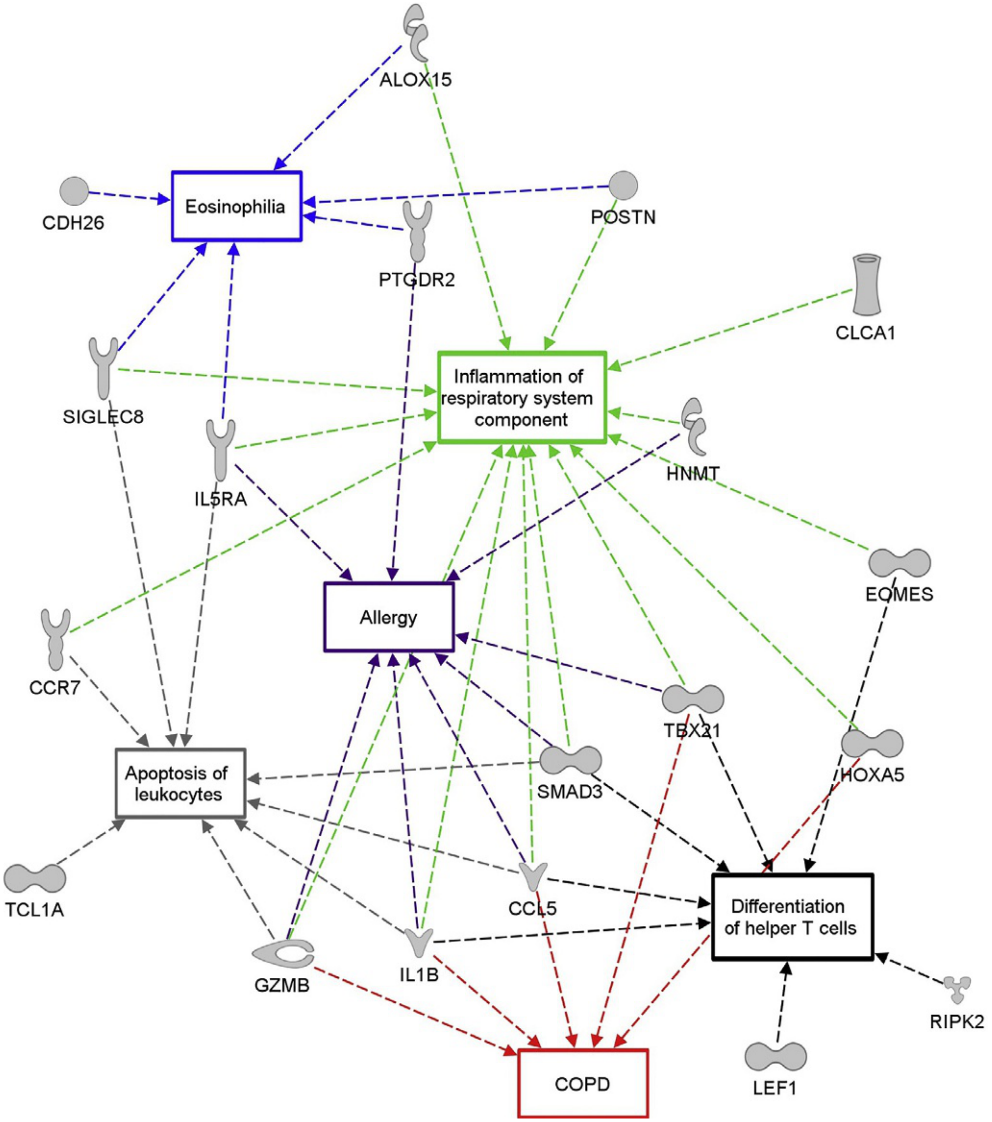

ALOX15
CDH26
Eosinophilia
POSTN
PTGDR2
CLCA1
SIGLEC8
Inflammation of respiratory system component
IL5RA
HNMT
Allergy
EOMES
CCR7
TBX21
Apoptosis of leukocytes
SMAD3
HOXA5
TCL1A
CCL5
Differentiation of helper T cells
GZMB
IL1B
RIPK2
COPD
LEF1

Fig. 1 Pathways associated with differential methylation. Genes and gene products associated with differential methylation discussed in this review are shown as gray symbols and selected pathways (shown as rectangles) have been added to the figure. The arrows indicate which genes are members of each pathway.

than differences—a goal that is put to a severe test when study populations are profoundly heterogeneous in their phenotypic characterization, age, geographic location (and therefore environment), sampling strategies, and sometimes even ethnicity. This problem may be especially acute for childhood asthma, the diagnostic criteria for which are not uniformly defined. Perhaps not by chance, epigenome-wide DNA methylation surveys provide more robust results for allergic sensitization, which is a more objective, quantitative, and discriminating phenotype. Obviously, carefully characterized longitudinal mother-child cohorts would be optimal for this type of studies, but are not available. Because the populations currently accessible are unevenly phenotyped and highly heterogeneous in several critical respects, more focused analyses may be more appropriate.

Technical issues may also undermine our ability to decipher the epigenetic determinants of allergic diseases. Epigenome-wide surveys are not truly epigenome wide, and the platform on which virtually all studies have relied so far may well be less than ideal: key regulatory loci might go unexplored, or coverage of these loci might be too sparse. A platform that expands coverage of putative regulatory elements has recently become available, but only a few studies have used it so far. Whole-genome bisulfite sequencing might overcome these technical limitations, but costs are still high and more effective analytical pipelines need to be developed. Finally, it has been recently proposed that weak *trans*-effects on gene expression may account for the largest proportion of complex trait heritability [49]. Because most analyses have correlated differential methylation and gene expression over relatively short genomic distances (5–10 kb), their results may have to be revisited.

In conclusion, DNA methylation studies in asthma and allergies have already provided interesting and sometimes tantalizing evidence about the time of inception of these conditions, their trajectories, and candidate pathways involved in established disease. We expect that the full potential of allergic disease epigenetics will be revealed once technical advances will allow unbiased, truly epigenome-wide surveys in newly designed longitudinal cohorts well characterized in their disease phenotypes and environmental exposures.

References

[1] Halonen M, et al. Perinatal TNF-alpha production influenced by maternal pregnancy weight gain, predicts childhood asthma. Am J Respir Crit Care Med 2013;188(1):35–41.

[2] Martinez FD, Vercelli D. Asthma. Lancet 2013;382(9901):1360–72.

[3] Rothers J, et al. Adaptive cytokine production in early life differentially predicts total IgE levels and asthma through age 5 years. J Allergy Clin Immunol 2011;128(2):397–402.

[4] Moore WC, et al. Identification of asthma phenotypes using cluster analysis in the Severe Asthma Research Program. Am J Respir Crit Care Med 2010;181(4):315–23.

[5] Wenzel SE. Asthma phenotypes: the evolution from clinical to molecular approaches. Nat Med 2012;18(5):716–25.

[6] Meyers DA, et al. Asthma genetics and personalised medicine. Lancet Respir Med 2014;2(5):405–15.

[7] Pividori M, et al. Shared and distinct genetic risk factors for childhood-onset and adult-onset asthma: genome-wide and transcriptome-wide studies. Lancet Respir Med 2019;7(6):509–22.

[8] Manolio TA. Genomewide association studies and assessment of the risk of disease. N Engl J Med 2010;363(2):166–76.

[9] Baurecht H, et al. Genome-wide comparative analysis of atopic dermatitis and psoriasis gives insight into opposing genetic mechanisms. Am J Hum Genet 2015;96(1):104–20.

[10] Ober C. Asthma genetics in the post-GWAS era. Ann Am Thorac Soc 2016;13(Suppl 1):S85–90.

[11] Toskala E, Kennedy DW. Asthma risk factors. Int Forum Allergy Rhinol 2015;5(Suppl 1):S11–6.
[12] Beasley R, Semprini A, Mitchell EA. Risk factors for asthma: is prevention possible? Lancet 2015; 386(9998):1075–85.
[13] Schwartz J, et al. Predictors of asthma and persistent wheeze in a national sample of children in the United States. Association with social class, perinatal events, and race. Am Rev Respir Dis 1990;142(3):555–62.
[14] Douwes J, et al. Farm exposure in utero may protect against asthma, hay fever and eczema. Eur Respir J 2008;32(3):603–11.
[15] DeVries A, Vercelli D. Epigenetics of human asthma and allergy: promises to keep. Asian Pac J Allergy Immunol 2013;31(3):183–9.
[16] DeVries A, Vercelli D. Epigenetic mechanisms in asthma. Ann Am Thorac Soc 2016;13(Suppl 1): S48–50.
[17] DeVries A, Vercelli D. Early predictors of asthma and allergy in children: the role of epigenetics. Curr Opin Allergy Clin Immunol 2015;15(5):435–9.
[18] DeVries A, et al. Epigenome-wide analysis links SMAD3 methylation at birth to asthma in children of asthmatic mothers. J Allergy Clin Immunol 2017;140(2):534–42.
[19] den Dekker HT, et al. Newborn DNA-methylation, childhood lung function, and the risks of asthma and COPD across the life course. Eur Respir J 2019;53.
[20] Reese SE, et al. Epigenome-wide meta-analysis of DNA methylation and childhood asthma. J Allergy Clin Immunol 2019;143:2062–74.
[21] Peng C, et al. Epigenetic age acceleration is associated with allergy and asthma in children in Project Viva. J Allergy Clin Immunol 2019;143:2263–70.e14.
[22] Moffatt MF, et al. A large-scale, consortium-based genomewide association study of asthma. N Engl J Med 2010;363(13):1211–21.
[23] Ferreira MA, et al. Genome-wide association analysis identifies 11 risk variants associated with the asthma with hay fever phenotype. J Allergy Clin Immunol 2014;133(6):1564–71.
[24] Hinds DA, et al. A genome-wide association meta-analysis of self-reported allergy identifies shared and allergy-specific susceptibility loci. Nat Genet 2013;45(8):907–11.
[25] Hastie AT, et al. Analyses of asthma severity phenotypes and inflammatory proteins in subjects stratified by sputum granulocytes. J Allergy Clin Immunol 2010;125(5):1028–1036.e13.
[26] Raedler D, et al. Identification of novel immune phenotypes for allergic and nonallergic childhood asthma. J Allergy Clin Immunol 2014;135(1):81–91.
[27] Simpson JL, et al. Elevated expression of the NLRP3 inflammasome in neutrophilic asthma. Eur Respir J 2014;43(4):1067–76.
[28] Anthoni M, et al. Transforming growth factor-beta/Smad3 signalling regulates inflammatory responses in a murine model of contact hypersensitivity. Br J Dermatol 2008;159(3):546–54.
[29] Bartram U, Speer CP. The role of transforming growth factor beta in lung development and disease. Chest 2004;125(2):754–65.
[30] Basu R, et al. IL-1 signaling modulates activation of STAT transcription factors to antagonize retinoic acid signaling and control the TH17 cell-iTreg cell balance. Nat Immunol 2015;16(3):286–95.
[31] Yang IV, et al. DNA methylation and childhood asthma in the inner city. J Allergy Clin Immunol 2015;136(1):69–80.
[32] Marioni RE, et al. The epigenetic clock is correlated with physical and cognitive fitness in the Lothian Birth Cohort 1936. Int J Epidemiol 2015;44(4):1388–96.
[33] Marioni RE, et al. DNA methylation age of blood predicts all-cause mortality in later life. Genome Biol 2015;16:25.
[34] Levine ME, et al. DNA methylation age of blood predicts future onset of lung cancer in the women's health initiative. Aging (Albany NY) 2015;7(9):690–700.
[35] Christiansen L, et al. DNA methylation age is associated with mortality in a longitudinal Danish twin study. Aging Cell 2016;15(1):149–54.
[36] Chen BH, et al. DNA methylation-based measures of biological age: meta-analysis predicting time to death. Aging 2016;8(9):1844–65.
[37] Perna L, et al. Epigenetic age acceleration predicts cancer, cardiovascular, and all-cause mortality in a German case cohort. Clin Epigenetics 2016;8:64.

[38] Simpkin AJ, et al. Prenatal and early life influences on epigenetic age in children: a study of mother-offspring pairs from two cohort studies. Hum Mol Genet 2016;25(1):191–201.
[39] Simpkin AJ, et al. The epigenetic clock and physical development during childhood and adolescence: longitudinal analysis from a UK birth cohort. Int J Epidemiol 2017;46(2):549–58.
[40] Binder AM, et al. Faster ticking rate of the epigenetic clock is associated with faster pubertal development in girls. Epigenetics 2018;13(1):85–94.
[41] Horvath S. DNA methylation age of human tissues and cell types. Genome Biol 2013;14(10):R115.
[42] Knight AK, et al. An epigenetic clock for gestational age at birth based on blood methylation data. Genome Biol 2016;17(1):206.
[43] Edris A, et al. Epigenome wide association studies in asthma: a systematic review. Clin Exp Allergy 2019;49:953–68.
[44] Xu CJ, et al. DNA methylation in childhood asthma: an epigenome-wide meta-analysis. Lancet Respir Med 2018;6:379–88.
[45] Reinius LE, et al. Differential DNA methylation in purified human blood cells: implications for cell lineage and studies on disease susceptibility. PLoS One 2012;7(7).e41361.
[46] Zhang H, et al. DNA methylation and allergic sensitizations, a genome-scale longitudinal study during adolescence. Allergy 2019;74:1166–75.
[47] Yang IV, et al. The nasal methylome and childhood atopic asthma. J Allergy Clin Immunol 2017;139(5):1478–88.
[48] Forno E, et al. DNA methylation in nasal epithelium, atopy, and atopic asthma in children: a genome-wide study. Lancet Respir Med 2019;7(4):336–46.
[49] Liu X, Li YI, Pritchard JK. Trans effects on gene expression can drive omnigenic inheritance. Cell 2019;177(4):1022–34.e6.

This page intentionally left blank

CHAPTER 13

Epigenetic regulation of normal hematopoiesis and its dysregulation in hematopoietic malignancies

Mark Hartmann[a], Mariam Hakobyan[a,b], Jens Langstein[a,b], Maximilian Schönung[a,b], Sina Stäble[a,b], Aurore Touzart[c], Justyna A. Wierzbinska[b,c], Daniel B. Lipka[a]

[a]Section Translational Cancer Epigenomics, Division of Translational Medical Oncology, National Center of Tumor Diseases (NCT) & German Cancer Research Center (DKFZ), Heidelberg, Germany
[b]Faculty of Biosciences, Heidelberg University, Heidelberg, Germany
[c]Division of Cancer Epigenomics, German Cancer Research Center (DKFZ), Heidelberg, Germany

Contents

Epigenetic regulation of normal hematopoiesis 286
  Epigenetic modifications 286
  Cis-regulatory elements 287
  Normal hematopoiesis 287
  Epigenome dynamics during normal hematopoietic differentiation 288
Mutations of epigenetic regulators in clonal hematopoiesis and in hematopoietic neoplasms 290
  Preleukemic mutations 290
  Mutations in myeloid neoplasms 293
Epigenetic deregulation in T-cell acute lymphoblastic leukemia 297
  Leukemogenesis of T-ALL 297
  Alterations in epigenetic modifiers 298
  Methylation profiles in T-ALL 298
  Deregulation of the *TAL1* oncogene: A model for enhancer hijacking and oncogenic neo-enhancers 298
  Oncogenic neo-enhancers deregulating T-ALL oncogenes 299
Cell-of-origin identification using epigenetics: Chronic lymphocytic leukemia as a model 300
  Epigenetically deregulated genes as biomarkers in CLL 301
  Methylomics identify CLL subgroups of different cellular origins 302
  Identification of disease-specific DNA methylation patterns in CLL 303
Epigenomics for disease classification and as a diagnostic tool 304
  Epigenetic biomarkers 304
  DNA methylation biomarkers for cancer classification and risk assessment 305
  Computational models for cancer classification 306
Epigenetic therapies in hematopoietic neoplasms 307
  Inhibitors of epigenetic key players 307
  Mechanisms of epigenetic therapies and possible rationales for drug combinations 309
Acknowledgment 310
References 310

Epigenetics of the Immune System
https://doi.org/10.1016/B978-0-12-817964-2.00013-7

© 2020 Elsevier Inc.
All rights reserved.

Epigenetic regulation of normal hematopoiesis

Life-long production and regeneration of all mature blood cell types require a dynamic and tightly regulated process that guarantees an appropriate cell output, both during steady-state and under emergency conditions such as infection or bleeding. All mature blood cell types with their specialized functions and diverse phenotypes originate from a pool of multipotent hematopoietic stem cells. Epigenetic processes faithfully regulate lineage specification and differentiation programs and thereby ensure the functional and phenotypic diversity as well as appropriate production rates of all blood cell types [1].

Epigenetic modifications

Epigenetics investigates heritable phenotypes that are not mediated by changes in the DNA sequence. Epigenetic changes may result from external stimuli as well as from normal developmental and differentiation processes. At the molecular level, epigenetic changes affect gene expression patterns and their response to external stimuli. The epigenome of a cell consists of different regulatory layers, which include DNA methylation, histone modifications, nucleosome positioning, and the three-dimensional chromatin organization.

In mammalian cells, DNA methylation is mainly found at the C5 position of cytosines occurring in the context of CpG dinucleotides. 5-Methylcytosine (5mC) is probably the most investigated epigenetic mark so far and was first introduced as an epigenetic mechanism regulating X-chromosome silencing, genetic imprinting, and gene expression in 1975. Methylation of promoter regions typically correlates with gene silencing. DNA methyltransferases (DNMTs) are enzymes that transfer a methyl group to cytosines, while 10–11 translocation enzymes (TETs) catalyze the sequential oxidation of 5mC as part of the demethylation process. Because DNA methylation is a stable molecular mark that is able to mediate a heritable epigenetic memory, DNA methylation can be exploited to dissect differentiation trajectories and to identify the cell of origin in the context of malignant transformation [2].

At the chromatin level, histone modifications influence the chromatin structure and binding of chromatin remodeling enzymes. Histone modifications are covalent post-translational modifications of histone tails, including methylation, phosphorylation, and acetylation. Specific histone modifications, like acetylation of histone H3 lysine 27 (H3K27ac), are characteristic for active *cis*-regulatory regions in the genome, such as enhancers or promoters. In addition, nucleosome occupancy has a significant function for the accessibility of transcription factor binding sites and for components of the transcription machinery [2].

The spatial chromatin organization within the nucleus is relevant for interactions between cell type-specific *cis*-regulatory elements with gene promoters that control gene expression. Genome-wide chromosome conformation capture studies identified megabase-sized topologically associating domains (TADs), defined as strongly interacting chromatin within TAD boundaries. TADs seem to be highly conserved between cell types, however, intra-TAD interactions are very dynamic during differentiation processes [2].

Cis-regulatory elements

Gene expression is regulated by *cis*-regulatory elements, such as promoters and enhancers. Enhancers can be located in variable distance upstream or downstream of the promoter and usually contain specific transcription factor binding sites. The activity of enhancer elements is highly cell type-specific and enhancers play a key role in regulating cell type-specific gene expression patterns. Enhancers interact with promoters through chromatin loops that bring enhancers and promoters in close proximity to trigger the activation of the transcription machinery. This interaction is highly dynamic and mainly occurs within TADs. *Cis*-regulatory regions are defined by characteristic epigenetic traits. Whole-genome bisulfite sequencing, chromatin immunoprecipitation sequencing (ChIP-seq), and assay for transposase-accessible chromatin using sequencing (ATAC-seq) enabled genome-wide mapping of *cis*-regulatory elements and the characterization of key transcription factor binding patterns in a wide variety of cell types across different organisms. Typically, DNA methylation in the promoter and enhancer regions negatively correlates with gene expression. In addition, promoters and enhancers are devoid of nucleosomes to allow the binding of transcription factors. Adjacent nucleosomes are marked with characteristic histone modifications. Histone H3 lysine 4 trimethylation (H3K4me3) and lysine 27 acetylation (H3K27ac) are found at active promoter regions, while Histone H3 lysine 4 monomethylation (H3K4me1) and lysine 27 acetylation (H3K27ac) are associated with active enhancer regions [2].

Normal hematopoiesis

Hematopoietic stem cells (HSCs) are multipotent cells, which possess self-renewal capacity and are responsible for the life-long production of all mature blood and immune cell types. Under homeostatic conditions, HSCs reside in a quiescent state in the bone marrow but can exit dormancy to contribute to blood production in response to infection, inflammation, or blood loss.

The classical hierarchical model of hematopoiesis postulates that HSCs sit at the apex of a tree-like bifurcation model and produce progenitor cell populations in a stepwise differentiation process accompanied by a gradual loss of self-renewal capacity. HSCs produce multipotent progenitor cells (MPPs), which are still able to give rise to all blood and immune cell types, but have reduced self-renewal capacity. MPPs, in turn, differentiate into oligopotent progenitor cells (e.g., common myeloid progenitors, CMPs; common lymphoid progenitors, CLPs) with restricted differentiation potential for either the myelo-erythroid or the lymphoid lineages. Oligopotent progenitor cells further produce unipotent precursor cell populations and ultimately differentiated blood cells. All these cell populations can be purified from the bone marrow by a panel of surface markers using fluorescence-activated cell sorting (FACS). According to the classical model of hematopoiesis, the different hematopoietic stem and progenitor cell populations are considered to be homogenous cell populations that differentiate further at defined branch points according to binary decisions (Fig. 1). Importantly, this hierarchical model of hematopoiesis was based on the functional characterization of surface marker-defined cell populations using ex vivo and in vivo assays to test lineage potential and self-renewal capacity [3].

Recent studies applying single-cell approaches have challenged the classical hierarchical model of hematopoiesis as they revealed an unexpected functional and molecular heterogeneity within defined hematopoietic stem and progenitor cell populations. Moreover, only a small fraction of immunophenotypically defined HSCs are actually able to produce multilineage progeny while the majority of HSCs have a more restricted uni- or bilineage differentiation potential. The observed lineage bias of immunophenotypic HSCs was stable in serial transplantation assays indicating a molecular priming of HSCs. It is noteworthy to mention that the existence of lineage primed HSCs has been confirmed using barcoding experiments in unperturbed hematopoiesis. The heterogeneity of hematopoietic cell populations was also shown at the level of oligopotent progenitors. Several studies demonstrated that oligopotent hematopoietic cell populations such as CMPs, granulocyte-macrophage progenitors (GMPs), and megakaryocyte-erythroid progenitors (MEPs) consist of a mixture of functionally committed progenitor subpopulations with unilineage potential. The revised model of hematopoiesis suggests that unilineage progenitors can emerge directly from HSCs and that they cumulatively and continuously acquire lineage-commitment without passing through discrete progenitor cell populations (Fig. 1) [1, 3].

Epigenome dynamics during normal hematopoietic differentiation

Hematopoiesis is a tightly regulated and highly dynamic process. Epigenetic integrity is essential for normal hematopoiesis, and defects in the epigenetic machinery are associated with hematopoietic abnormalities. One of the first experimental hints for a central role of DNA methylation in the regulation of hematopoietic differentiation came from Dnmt1 hypomorphic mice. Reduced activity of Dnmt1 leads to hypomethylation of myelo-erythroid genes in hematopoietic stem and progenitor cells (HSPCs) and thereby promoted myelo-erythroid differentiation at the expense of lymphoid differentiation [4]. Later studies investigating DNA methylation patterns in several cell populations across the hematopoietic hierarchy demonstrated the high plasticity of the DNA methylome during hematopoietic differentiation and identified DNA methylation patterns associated with lineage choice and specific terminally differentiated blood cell types. Systematic investigation of

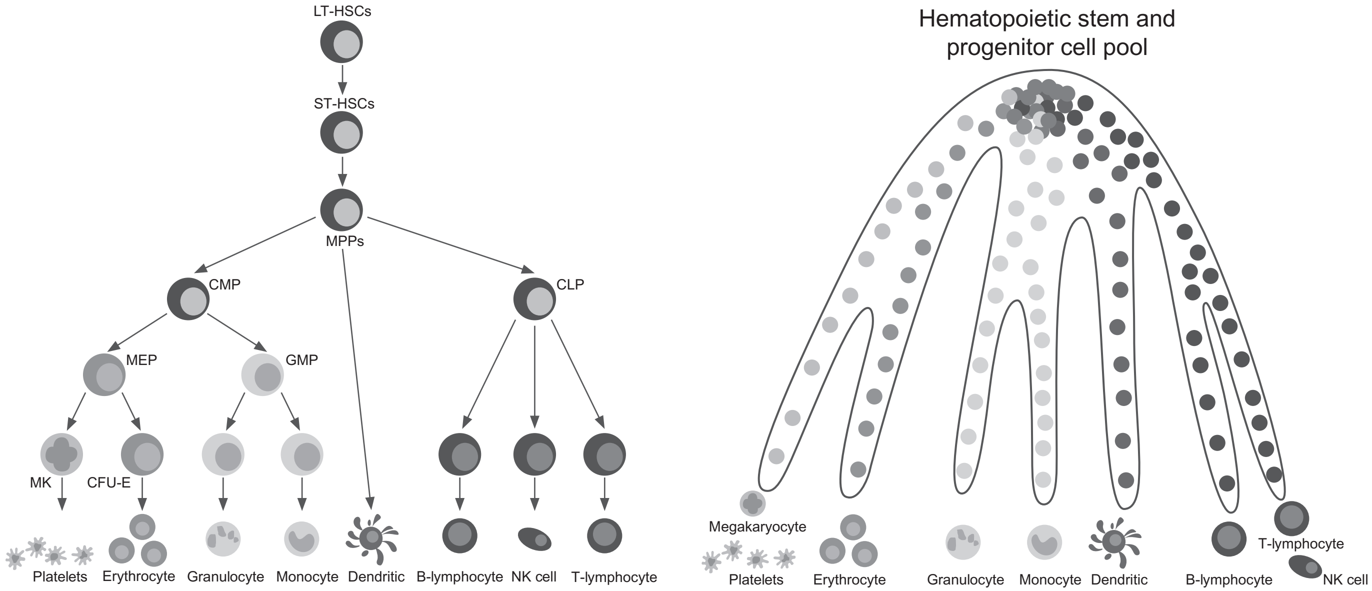

LT-HSCs
ST-HSCs
MPPs
CMP
CLP
MEP
GMP
MK
CFU-E
Platelets
Erythrocyte
Granulocyte
Monocyte
Dendritic
B-lymphocyte
NK cell
T-lymphocyte
Hematopoietic stem and progenitor cell pool
Megakaryocyte
Platelets
Erythrocyte
Granulocyte
Monocyte
Dendritic
B-lymphocyte
T-lymphocyte
NK cell

Fig. 1 Main concepts of normal hematopoiesis: the classical hierarchical model of discrete cell types and the contemporary continuum model of indefinite cell states. *(Modified from Laurenti E, Gottgens B. From haematopoietic stem cells to complex differentiation landscapes. Nature 2018;553 (7689):418–26).*

differentially methylated regions (DMRs) revealed an enrichment of lineage-specific transcription factor binding sites and linked DNA methylation dynamics to gene expression changes. These findings led to the identification of novel factors that regulate lymphoid versus myeloid differentiation and further established DNA methylation to function as an epigenetic switch that determines differentiation capacity and cell identity by preventing aberrant transcription factor binding [5–7]. A genome-wide DNA methylation map of the murine HSPC compartment demonstrated that widespread DNA methylation changes occur already during early hematopoietic commitment. Dynamic DNA methylation changes were found predominantly in distant *cis*-regulatory regions, affecting both the expression of stemness-associated and lineage-specific genes [6]. Finally, a comprehensive genome-wide DNA methylation map of human blood cells that were isolated from different tissues, such as bone marrow and peripheral blood, revealed variations in their DNA methylation profiles depending on their tissue source [7].

Similar to DNA methylation, genome-wide chromatin accessibility studies using ATAC-seq revealed open chromatin regions that are highly cell type-specific and enriched for binding motifs of key hematopoietic transcription factors [8]. De novo enhancers as defined by ChIP-seq demonstrated an increase of chromatin accessibility and lineage-specific enhancer regions were enriched for lineage-specific transcription factor binding motifs [9]. This suggests that enhancer dynamics precede transcriptional programs during hematopoietic differentiation. Interestingly, chromatin accessibility has better deconvoluted cell populations compared to gene expression, emphasizing that epigenetic programming reflects cell identity much better than transient transcriptional states.

Collectively, the current knowledge of epigenome dynamics during hematopoietic differentiation demonstrates that the epigenome reflects and likely instructs hematopoietic lineage commitment and differentiation trajectories. Various studies suggest that epigenetic patterns are considerably precise in defining cell identity and differentiation stages, indicating that the epigenetic landscape provides the framework for the possible transcriptional responses. This also implies that the epigenomic patterns are established before gene expression programs are initiated. Future work integrating newly developed single-cell methods for epigenomics, transcriptomics, proteomics, and lineage tracing will provide unprecedented molecular insights into how hematopoietic differentiation is regulated in individual cells at the molecular level.

Mutations of epigenetic regulators in clonal hematopoiesis and in hematopoietic neoplasms

Preleukemic mutations

In human adults, about 50,000–200,000 HSCs are estimated to actively contribute to blood production at any given time [10]. Throughout the lifetime of an individual, these cells accumulate roughly 0.13 exonic mutations per year, suggesting that the HSC population encompasses a mosaic of individual clones where each carries a distinct mutational signature [11]. While the majority of these mutations will have a neutral effect, occasionally these somatic events will confer a competitive advantage to this particular HSC, thereby promoting the expansion of this clone at a disproportionate rate compared to other clones [12]. The frequency of clonal expansion in the hematopoietic system strongly correlates with age and can be found in otherwise healthy individuals that do not show any other hematopoietic abnormalities. Over recent years, several studies identified recurrent mutational patterns that were associated with clonal hematopoiesis, often comprising known driver genes of hematopoietic malignancies. Clonal hematopoiesis is currently not considered as a disease, as the associated risk of malignant transformation is low (about 1% per year). Nevertheless, individuals with clonal hematopoiesis display a ~10-fold increased risk of eventually developing a hematopoietic neoplasm, suggesting that these somatic mutations give rise to a preneoplastic or preleukemic state. Such a preleukemic state evolutionarily precedes the development of hematopoietic neoplasms and can be detected years before the diagnosis of overt disease [13].

Clonal hematopoiesis may arise simply through aging-related stochastic processes, which impose cell-intrinsic selective pressure and which are coupled to the functional and molecular properties of aged HSCs. Compared to younger counterparts, older HSCs display a decline in self-renewal capacity and regenerative potential but also show a differentiation bias towards the myeloid lineage [14]. At the molecular level, this functional decline is associated with the proliferative history of HSCs. This correlates with an age-dependent increase in DNA damage, which accumulates over time in response to replicative and oxidative stress, radiation, or toxicity. Similarly, it has been demonstrated that HSC functionality highly depends on epigenomic integrity as the functional decline of HSCs correlates with an aberrant epigenetic landscape. Depending on the proliferative history, chronologically aged HSCs exhibit an increase in DNA methylation levels at sites enriched for differentiation-promoting genes and a simultaneous decrease at HSC maintenance genes. The importance of epigenetic integrity for HSC functionality is further underlined by the observation that genes-encoding epigenetic regulators are the most frequently mutated gene category in acute myeloid leukemia (AML) and these mutations are often early events during AML pathogenesis based on their variant allele frequency. Strikingly, the same epigenetic regulators, which are frequently mutated in AML, such as *DNMT3A*, *TET2*, and *ASXL1*, are also frequently mutated in individuals with clonal hematopoiesis.

DNMT3A encodes a de novo DNA methyltransferase, which establishes DNA methylation marks at previously unmethylated CpG dinucleotides. The majority of *DNMT3A* mutations affect the methyltransferase domain, with an enrichment of missense mutations found at codon R882 (65% of all *DNMT3A* mutations) [15]. These mutations impair the catalytic activity of DNMT3A, and consequently, R882 mutations result in specific DNA methylation patterns that are dominated by local hypomethylation. Phenotypically, mouse models expressing the DNMT3A-R882H mutation or the murine homolog Dnmt3a-R778H demonstrate a mild phenotype with a discreet expansion of hematopoietic stem and progenitor cells in the bone marrow, thereby recapitulating clonal hematopoiesis [16]. Although hypomethylation is considered an initiating event for the development of clonal hematopoiesis and its progression to AML, spontaneous malignant transformation is only rarely if at all observed in the presence of the *DNMT3A* mutations alone. It is believed that *DNMT3A* mutations provide a fertile ground for malignant transformation and maybe even for the acquisition of secondary mutations. Along these lines, it has been shown that *DNMT3A* mutations synergize with other mutations as discussed below.

TET2 encodes a member of the methylcytosine dioxygenase class, which catalyzes the oxidation of 5mC. Loss-of-function mutations of *TET2* are often found in clonal hematopoiesis. This results in widespread DNA hypermethylation, which is believed to impact on the regulation of close-by genes. *TET2* mutations are associated with an expansion of the hematopoietic stem and progenitor compartment and an increased HSC self-renewal [17]. However, similar to *DNMT3A* mutations, *TET2* mutations alone only rarely lead to the development of hematopoietic neoplasms. Yet, data demonstrating hypermutagenicity upon *TET2*-loss and synergism with other oncogenic driver mutations such as *FLT3*-ITD and *JAK2*-V617F illustrate the potential of *TET2* mutations to drive the progression from clonal hematopoiesis to overt hematopoietic neoplasms.

Mutations in the metabolic enzymes IDH1 and IDH2 almost exclusively affect the amino acid residues R132 or R172, respectively. *IDH1/2* mutations lead to neoenzymatic activity of IDH1/2 resulting in the production of the oncometabolite 2-hydroxyglutarate (2HG). 2HG is a competitive inhibitor of enzymes of the dioxygenase class including TET2. Studies investigating the clonal hierarchy in AML patients have detected *IDH1* and *IDH2* mutations at high-variant allele frequencies in residual normal HSCs, suggesting that *IDH1/2* mutations occur early during pathogenesis, likely even before the actual malignant transformation has occurred [18]. Interestingly, although *IDH1/2* mutations are early leukemogenic events, unlike *DNMT3A* and *TET2* mutations, *IDH1*/2 mutations are rarely found in individuals with clonal hematopoiesis. Several lines of evidence suggest that *IDH1/2* mutations might be associated with higher oncogenic potential than other mutations found in clonal hematopoiesis, but this is a matter of ongoing research. Because of their mutual exclusivity with *TET2* mutations, it was initially thought that the *IDH1/2* mutations mediate their oncogenic potential via the inhibition of TET2 enzymatic activity [19]. However, several TET2-independent mechanisms have been proposed that trace back to the complex and pleiotropic effects of 2HG within the cell. These mechanisms include, for example, dysbalance of reactive oxygen species, inhibition of histone demethylases, and downregulation of several genes involved in DNA repair.

ASXL1 encodes a protein that is involved in the regulation and recruitment of the PRC2 complex, which catalyzes the methylation of lysine 27 at histone H3. Mutations in *ASXL1* have been observed both in healthy individuals with clonal hematopoiesis and in patients with hematopoietic neoplasms such as myelodysplastic syndromes, chronic myelomonocytic leukemia, or AML. *ASXL1* mutations frequently result in loss-of-function of ASXL1 and subsequently to a loss of PRC2-mediated gene repression [20]. As a direct consequence, the derepression of PRC2 target genes leads to functional alterations of HSPCs and promotes myeloid transformation. However, neoplastic transformation in mice carrying an *ASXL1* mutation was observed only after long latencies, suggesting that additional factors are required for disease progression and malignant transformation.

In summary, preleukemic mutations frequently affect epigenetic regulators. Their aberrant activity results in epigenetic alterations and thus affects epigenomic integrity during hematopoiesis. While the functional link between the presence of preleukemic mutations and an increased risk of hematopoietic neoplasms is well established, it still remains a matter of ongoing research to decipher the underlying molecular mechanisms driving malignant transformation in the presence of preleukemic mutations affecting epigenetic regulators.

Mutations in myeloid neoplasms

Myeloid neoplasms such as AML, myeloproliferative neoplasms (MPN), myelodysplastic syndromes (MDS), and myelodysplastic/myeloproliferative neoplasms are clonal disorders of hematopoiesis.

AML is a highly heterogeneous disease originating from hematopoietic stem and progenitor cells. AML is defined by an abnormal, clonal proliferation of immature and dysfunctional cells, so-called leukemic blasts, which are characterized by the expression of myeloid markers. These leukemic blasts infiltrate the bone marrow, blood, and occasionally other tissues [21]. Typically, a diagnosis of AML requires the presence of $\geq$20% blasts in either peripheral blood or bone marrow. The uncontrolled accumulation of leukemic blasts in the bone marrow interferes with the production of normal blood, probably by cytokines and chemokines that are produced by those blasts [22]. This affects the proliferation and differentiation of residual healthy or preleukemic HSCs. AML accounts for 80% of all acute leukemia cases and can arise as a consequence of prior radio- or chemotherapy (referred to as therapy-related AML) or secondary to preleukemic states such as MDS (secondary AML). Most cases, however, occur in previously healthy individuals (de novo AML). AML is mainly a disease of the elderly with a mean age at diagnosis of 70 years, although it can also affect children and younger adults. While the patients with pediatric AML mostly display very low mutation frequencies, AML in older individuals is typically characterized by the presence of preleukemic mutations in combination with additional leukemic driver mutations.

AML is characterized by a multitude of genetic and genomic aberrations. Somatic mutations can be detected in leukemic blasts of almost all AML patients. Overall, more than 70% of all AML patients harbor mutations affecting genes that encode an epigenetic regulator. The most frequently mutated epigenetic regulator is DNMT3A, which is found in 20%–30% of all AML patients [15]. *DNMT3A* mutations are often found in combination with mutations of *FLT3*, *NPM1*, and *IDH1*. Other frequent mutations of epigenetic regulators affect TET2 and ASXL1. As mentioned above, many of these epigenetic factors are often altered in both AML and clonal hematopoiesis and are therefore considered as early events during disease pathogenesis. It is believed that remodeling of the epigenetic landscape provides a fertile ground for the acquisition of second hit mutations (e.g., *FLT3* and *NPM1*) or at least for the outgrowth of cells that have acquired such a second hit mutation. This is thought to promote malignant transformation and to accelerate the development of the disease. However, the exact molecular and cellular mechanism of how epigenetic alterations contribute to leukemogenesis remains largely unknown and is a matter of intense research.

Recent studies have identified AML subclasses based on DNA methylation profiles. These subclasses largely reflect genetic or cytogenetic aberrations meaning that the DNA methylation signatures correlate with specific genetic lesions. Silencing of several proapoptotic and tumor suppressor genes, as well as the upregulation of proto-oncogenes by altered DNA methylation of their promoter regions, has been demonstrated in AML. In addition, there is data suggesting that mutations in epigenetic regulators lead to genetic instability and consequently predispose to the acquisition of additional mutations [23].

While most cases of AML are diagnosed in older adults and are associated with a typical pattern of co-occurring mutations, pediatric AML cases are mostly characterized by chromosomal aberrations. Generally, cytogenetic heterogeneity in AML has been recognized over 30 years ago. Overall, about 50%–60% of all newly diagnosed AML patients harbor at least one of the >700 known chromosomal aberrations that have been cataloged in AML [24]. The most common karyotypic abnormality in AML is the translocation t(8;21), while translocation t(15;17) is pathognomonic for acute promyelocytic leukemia (APL). Each of these translocations is found in up to 10% of all cases, and result in the fusion-proteins RUNX1-RUNX1T1 and PML-RARA, respectively. The third most common abnormality is the inversion of chromosome 16, inv(16), which occurs in 5% of AML cases with cytogenetic abnormalities and results in the CBFB-MYH11 fusion protein. Such rearrangements can co-occur with altered epigenetic profiles and might affect the prognosis of patients.

As mentioned above, infant and pediatric leukemias are often characterized by chromosomal aberrations, which in some cases already arise in utero during fetal hematopoiesis. Up to 80% of all t(4;11) translocations result in the expression of the fusion transcripts MLL-AF4 and AF4-MLL, respectively. The wild-type gene *mixed-lineage leukemia*

(*MLL*) or *KMT2A* encodes for an H3K4me3 histone methyltransferase, which plays a central role in maintaining gene activation during normal embryonic development [25]. MLL enables the expression of genes such as homeobox family members, which play a central role in the regulation of hematopoietic stem cell proliferation and differentiation. Aberrant activation of MLL has been shown to be involved in different types of leukemia including B-cell acute lymphoblastic leukemia, acute myeloid leukemia, and acute lymphoid leukemia. MLL can be dysregulated in different ways, including partial tandem duplication, gene amplification, internal deletions, and chromosomal translocations. The most frequent aberrations known to activate MLL are fusion transcripts. The MLL fusion partner can have an impact on leukemogenesis, and in the course of the disease, most MLL-driven leukemias are characterized by aberrant methylation of H3K79 [26]. H3K79 methylation is a histone mark that strongly correlates with transcriptional activity. While MLL itself targets H3K4, its fusion products are able to mistargeting the H3K79 methyltransferase DOT1L to MLL target genes. Mislocated DOT1L activity is directly increased by MLL fusion proteins and it is crucial for their leukemogenic potential.

In addition to chromosomal translocations resulting in fusion proteins, loss of chromosomal material via deletions or monosomies is frequently observed in AML. Monosomy of chromosome 7 or loss of genetic information from the long arm of chromosome 7, del(7q), occurs in about 10% of all MDS and AML cases but are most frequently detected in patients with complex karyotype AML [27]. To date, it remains unclear how mechanistically loss of material from chromosome 7 contributes to AML pathogenesis or to AML progression. Several candidate tumor suppressor genes have been proposed on chromosome 7, but none of the candidate genes have been shown to fully recapitulate features of AML with monosomy 7 or del(7q). Hence, the current working hypothesis is that several (haploinsufficient) tumor suppressor genes contribute to the −7/del(7q) AML phenotype in an additive or synergistic fashion.

Interstitial deletions of the long arm of chromosome 5, del(5q), contribute to a rare form of MDS, also known as 5q− syndrome. It is defined by a bone marrow blast count <5% and exhibits a low risk for leukemic transformation. While several distinct breakpoints on chromosome 5 have been detected, the most frequent deletions include del(5)(q13q33), del(5)(q13q31), and del(5)(q22q33). Due to the large variability of the breakpoints, the emergence of a novel oncogene is less probable. Therefore, most of the molecular research on 5q− syndrome has focused on the identification of a putative tumor suppressor gene. While 5q− syndrome is generally associated with low-risk MDS and good prognosis, del(5q) has also been associated with more aggressive phenotypes, especially when detected in the context of other cytogenetic abnormalities. Particularly, patients with more aggressive types of MDS or (complex karyotype) AML harbor deletions on chromosome 5q, which are distinct from the deletions detected in 5q− syndrome.

Loss of chromosomal material from the short arm of chromosome 17, del(17p), is also frequently observed in AML, and often results in inactivation of TP53.

A distinct mechanism of how a chromosomal aberration can contribute to leukemogenesis has been demonstrated for inversions/translocations of chromosome 3 (inv(3)/t(3;3)). While cytogenetic abnormalities often result in the loss of tumor suppressor genes or in fusion transcripts with oncogenic potential, inv(3)/t(3;3) is associated with the relocation of the *GATA2* enhancer. Normally, the *GATA2* enhancer regulates the expression of the hematopoietic transcription factor GATA2. Its relocation in inv(3)/t(3;3) patients results in aberrant expression of the oncogene *EVI1*, whose activation plays a central role in the pathology of inv(3)/t(3;3) AML. This phenomenon is called "enhancer hijacking." Additionally to the *EVI1* activation, the loss of the GATA2 enhancer from its original genetic locus results in reduced expression of the haploinsufficient transcription factor GATA2, which contributes to the phenotype of this enhancer-hijacking event.

Genomic amplifications represent an additional but rare mechanism for genetic/cytogenetic abnormalities observed in AML. Genomic amplifications with malignant phenotypes often affect oncogenes that are upregulated as a consequence of copy number gains. Chromosomal amplifications are known to play a role in the pathology of myeloid malignancies. The gain of chromosome 8 and chromosome 21 resulting in trisomy is the two most frequent events of this type in AML. In addition to trisomy 21, partial amplification of chromosome 21 have been observed. Patients with Down syndrome, harboring a germline trisomy 21, show a 10- to 20-fold increased risk for the development of myeloid malignancies compared to individuals without Down syndrome. This indicates a gene-dosage effect of genes located on chromosome 21. Similar mechanisms probably contribute to the leukemogenic potential of trisomy 8. A prominent gene located on chromosome 8 and known to exhibit gene-dosage effects is *c-MYC*, which is involved in the pathogenesis of several malignant disease entities.

While the aforementioned cytogenetic aberrations affect only one or at the maximum two chromosomes, complex aberrant karyotype AML (ck-AML) is characterized by multiple unrelated cytogenetic abnormalities [28]. Up to 12% of unselected AML patients and up to 23% of elderly AML patients harbor complex karyotype abnormalities, which are commonly defined as the presence of at least three cytogenetic alterations. Complex karyotype AML is associated with extremely poor prognosis. The most common abnormalities found in ck-AML include chromosomal deletions from 5q, 7q, and 17p. Up to 85% of all ck-AML patients display at least one of these chromosomal aberrations. Therefore, their presence characterizes the "typical" ck-AML patient. To date, ck-AML is only poorly characterized at the molecular level despite being one of the most frequent cytogenetic subgroups of AML.

Epigenetic deregulation in T-cell acute lymphoblastic leukemia

T-cell acute lymphoblastic leukemia (T-ALL) is a rare lymphoid neoplasm characterized by the proliferation of immature T-cell precursors arrested at specific stages of maturation. T-ALL is mainly diagnosed in adolescents and young adults (10–39 years) accounting for about 20% of acute lymphoblastic leukemia diagnosed in this age group. Using intensive chemotherapy, prognosis has improved significantly, especially in children where an overall survival rate of $>80\%$ can be achieved. In adult patients, overall survival reaches only about 60% due to a high rate of relapse.

Leukemogenesis of T-ALL

Genetically, T-ALL is a heterogeneous disease that results from the malignant transformation of immature thymic precursors [29, 30]. These cells accumulate genetic alterations that lead to activation of oncogenes, silencing of tumor suppressor genes, and disruption of pathways that are involved in the control of cell proliferation, survival, and thymocyte differentiation. Pathogenesis of T-ALL is a multistep process that involves several genetic and epigenetic events, which drive leukemogenesis by increasing the proliferation rate and by reducing the differentiation capacity of thymic precursors.

Chromosomal translocations involving the TCR loci ($TCR\alpha\delta$ and $TCR\beta$) are a feature of T-ALL and are observed in nearly 30% of cases. The translocation of a TCR *cis*-regulatory element into the vicinity of an oncogene is generally considered to be directly responsible for its ectopic expression. The most frequent translocation partners are genes-encoding transcription factors, including those of the bHLH family *TAL1*, *TAL2*, *LYL1*, and *BHLHB1*; those of the "LIM-only domain" family *LMO1* and *LMO2*; the homeobox genes *TLX1* (*HOX11*) and *TLX3* (*HOX11L2*); *HOXA*, *NKX2-1*, *NKX2-2* as well as the transcription factor genes *MYB* and *MYC*.

Gene expression profiling studies have identified several molecular subgroups of T-ALL that are associated with specific driver oncogenes: TAL/LMO (mature or late cortical stage), TLX1, TLX3, and NKX2-1/NKX2-2 (early cortical stage), LYL1/MEF2C and HOXA (immature stage). These transcription factors regulate distinct, mutually exclusive molecular pathways and are therefore often associated with particular stages of T-cell differentiation.

In addition to these subtype-defining alterations, many other recurrent genetic anomalies have been observed. This includes NOTCH1 pathway activation and biallelic deletion or epigenetic silencing of *CDKN2A*, both of which are found in more than 50% of T-ALL cases. Mutations or copy number alterations have been found in genes involved in signal transduction (e.g., *RAS*, *PI3K*, *PTEN*, *IL7R*, *JAK1*, JAK3, and *STAT5B*), epigenetic regulation (*DNMT3A*, *PHF6*, *SUZ12*, and *EZH2*), protein translation (e.g., *RPL10* and *CNOT3*), and in tumor suppressor genes (e.g., *LEF1* and *ETV6*).

Alterations in epigenetic modifiers

T-ALL is among the tumors with the highest frequency of mutations in genes encoding for epigenetic regulators. Sequencing studies have identified recurrent somatic alterations in T-ALL that affected several genes involved in regulating DNA methylation and post-translational histone modifications. This suggested that the deregulation of epigenetic homeostasis plays an important role in the malignant transformation process [31]. *DNMT3A* mutations have been identified in up to 18% of T-ALL patients, are mainly associated with an immature phenotype, and are associated with poor prognosis.

DNMT3A is mutated in about 18% of T-ALL patients, predominantly in adult cases and in immature T-ALL subtypes. Similar to AML, *DNMT3A* mutations are associated with adverse outcomes in adult T-ALL. Mutations in *IDH1* or *IDH2* are exclusively detected in adult T-ALL cases (2%–9%), and *TET1* mutations are found in about 6%–13% of T-ALL patients. The prognostic impact of *IDH1/2* or *TET1* mutations has not been conclusively studied yet.

The polycomb repression complex 2 (PRC2) is involved in transcriptional repression through the deposition of the repressive histone mark H3K27me3. Up to 25% of T-ALL patients show disrupted PRC2 function, which is mediated by loss of function mutations or by deletions of one of its core components *EZH2*, *SUZ12*, or *EED*. Despite the high frequency of mutations in epigenetic regulators found in T-ALL, similar to the situation in myeloid neoplasms, mechanistic insights into how these mutations contribute to the malignant transformation process are still lacking.

Methylation profiles in T-ALL

Surprisingly, only a few studies have systematically analyzed DNA methylomes in T-ALL. To date, studies investigating the DNA methylomes of adult patients are lacking. In pediatric patients, however, two T-ALL subgroups have been identified based on DNA methylation array analysis. One subgroup displays pronounced CpG island hypermethylation, the so-called CpG island methylator phenotype (CIMP-positive) and is associated with a good prognosis. In contrast, the CIMP-negative subgroup is enriched in TAL1 dysregulated cases and it is associated with an unfavorable outcome. In the future, it will be interesting to see whether mutations in epigenetic regulators are associated with specific DNA methylation patterns in adult T-ALL.

Deregulation of the *TAL1* oncogene: A model for enhancer hijacking and oncogenic neo-enhancers

TAL1 is one of the most frequently dysregulated oncogenes in T-ALL with about 30%–50% of patients showing *TAL1* overexpression. Under physiological conditions, TAL1 is a transcription factor that is essential for the development of hematopoiesis and of the vascular system. In definitive hematopoiesis, TAL1 is important for erythroid differentiation and it is normally stably repressed by epigenetic mechanisms during human thymopoiesis.

Events leading to aberrant expression of *TAL1* in the T-cell lineage are considered strong drivers of T-cell leukemogenesis [32, 33]. Aberrant *TAL1* expression results from two major genomic mechanisms. In 25% of T-ALL, the interstitial micro-deletion SIL-TAL1 can be found. The second mechanism is much rarer and involves the translocation t(1;14)(p32;q11) that brings *TAL1* under the control of the T-cell receptor alpha locus. In both cases, these chromosomal alterations change *cis*-regulatory elements resulting in the "hijacking" of enhancers active in normal T cells leading to an aberrant monoallelic *TAL1* expression.

In about 50% of TAL1+ T-ALL cases, the aberrant expression of *TAL1* cannot be explained by any of the genomic alterations mentioned above. This suggested the presence of additional genetic or epigenetic mechanisms leading to *TAL1* overexpression in T-ALL. In such unresolved cases, the expression of *TAL1* may be both monoallelic and thus compatible with alterations in *cis*, or biallelic suggesting a mechanism acting in *trans*. The physiological repression of TAL1 in the T-cell lineage is mediated by a strong deposition of the repressive H3K27me3 histone mark at the *TAL1* locus. H3K27me3 deposition is markedly decreased in TAL1+ T-ALL. Recently, a somatic alteration consisting of a recurrent focal micro-insertion (2–18 bp) in a noncoding intergenic region 7 kb upstream of *TAL1* has been detected. These micro-insertions have in common that they always generate de novo MYB binding sites, which create an oncogenic neo-enhancer that is manifested by a switch from the repressive H3K27me3 mark to a deposition of the activating H3K27ac histone mark. It has been shown that MYB binding sites are required to establish the de novo enhancer activity in this region. A model suggests that MYB recruits the epigenetic modifier CBP, resulting in the enrichment of activating H3K27ac marks, which in turn allow the recruitment of other transcription factors belonging to the TAL1 complex including TAL1 itself. This finally leads to aberrant *TAL1* expression (Fig. 2). As a result, these micro-insertions are able to selectively induce monoallelic *TAL1* expression from the mutated allele. These recurrent micro-insertions explain about 20% of previously unexplained TAL1+ T-ALL cases.

Oncogenic neo-enhancers deregulating T-ALL oncogenes

The concept of somatically acquired neo-enhancer activity is also applicable to other oncogenes involved in T-ALL pathogenesis. Recent studies have reported oncogenic neo-enhancers for *LMO1* and *LMO2*, which are frequently co-regulated with *TAL1*. Similarly, overexpression of the *MYC* oncogene has been shown in some T-ALL with *NOTCH1* deregulation. MYC is indeed a critical oncogene in NOTCH1-induced leukemogenesis. A recurrent focal somatic duplication 1.43 Mb downstream of the *MYC* locus results in an oncogenic "super-enhancer." This regulatory element is activated by NOTCH1 and leads to the deregulation of *MYC* *via* a long-distance interaction with the *MYC* promoter.

These recent studies are important because they have established a new paradigm of genomic alterations in noncoding regions which lead to the creation of neo-enhancers

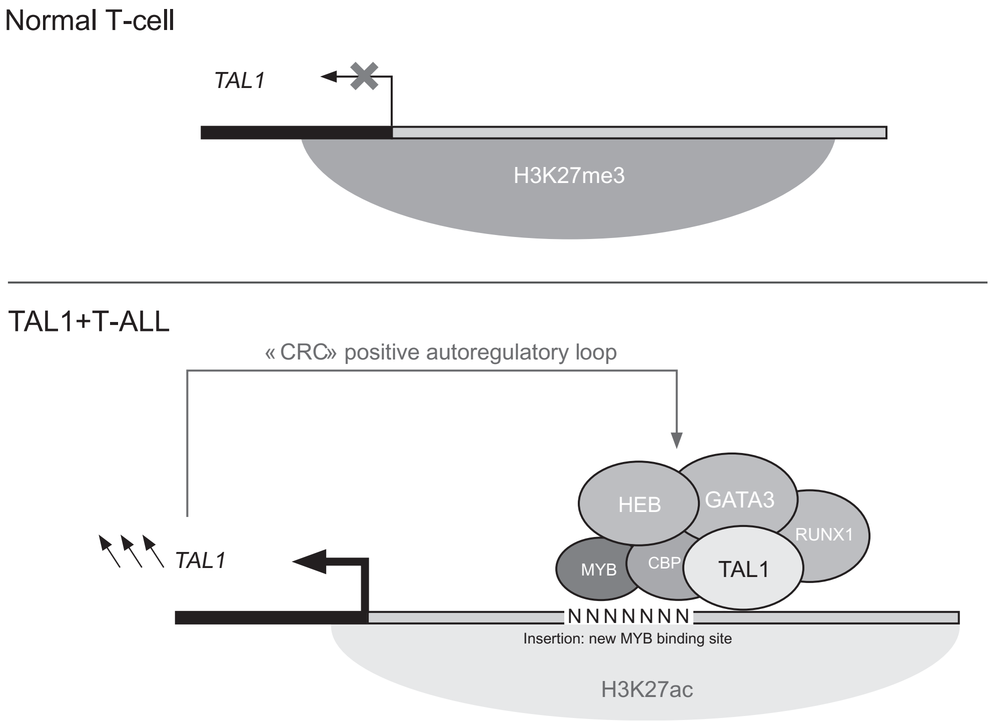

Normal T-cell
TAL1
H3K27me3
TAL1+T-ALL
« CRC» positive autoregulatory loop
HEB
GATA3
RUNX1
MYB
CBP
TAL1
TAL1
NNNNNNN
Insertion: new MYB binding site
H3K27ac

Fig. 2 Model for oncogenic neo-enhancers in T-ALL: a micro-insertion in a noncoding region upstream of the gene encoding the transcription factor TAL1 leads to epigenetic changes and the generation of a neo-enhancer that activates gene expression of *TAL1*. This gives rise to a positive autoregulatory loop and the transformation to T-ALL.

that activate the transcription of proto-oncogenes. This highlights how somatic mutations in noncoding intergenic regions can initiate oncogenesis. The generation of such oncogenic neo-enhancers potentially represents a general mechanism for oncogene activation. This underscores the importance of studying the functional impact of the noncoding genome. While intragenic mutations are well-studied, those mutations affecting noncoding regions have long been neglected. The development of high-throughput molecular biology techniques opens up the possibility to systematically study the functional impact of noncoding mutations.

Cell-of-origin identification using epigenetics: Chronic lymphocytic leukemia as a model

Aberrant epigenetic regulation such as altered DNA methylation patterns can be found in many cancer entities, including hematopoietic neoplasms. Malignant transformation can principally occur at any point during normal differentiation, which impacts on the epigenomic signature of the malignant cell. Major advances in our current understanding of aberrant DNA methylation programming come from studies investigating B-cell malignancies. Normal B-cell development is associated with broad DNA methylation dynamics, and this process is being disrupted in the pathogenesis of B-cell neoplasms. Insights gained through genome-wide DNA methylation profiling efforts in chronic lymphocytic leukemia (CLL), a malignancy arising from $CD19^+$ B cells, have demonstrated that different epigenetic signatures specify different CLL subgroups. Since most of the progress in understanding aberrant DNA methylation patterns in the context of normal differentiation has been made in the field of CLL, we will focus on CLL as a model disease in this section. We will discuss findings from gene-specific and genome-wide studies of DNA methylation in CLL and describe how to distinguish disease-specific aberrant DNA methylation from differentiation-related, normal DNA methylation. Such characterizations are crucial to identify the disease-specific cell of origin, which defines critical and individual biological features. This enables the identification of specific therapeutic targets, the exclusion of inefficient medication, and thus personalized therapies.

Epigenetically deregulated genes as biomarkers in CLL

Aberrant DNA methylation of individual genes can serve as a biomarker and help to identify genes that contribute to disease initiation or progression. Numerous studies in CLL describe genes that are deregulated by either DNA hypo- or hypermethylation. DNA hypomethylation of gene promoters can lead to the transcriptional activation of proto-oncogenes. When intergenic regions are hypomethylated this can lead to genomic instability. In contrast, DNA hypermethylation has been linked to tumor suppressor gene inactivation. In CLL, the list of aberrantly methylated genes is extensive and includes many protein-coding genes such as *TCL1*, *CHD1*, *DAPK1*, *ID4*, *ZAP70*, and *TWIST2*, as well as miRNAs such as *miR-21*, *miR-29a*, or *miR-34a* [34].

The first study of aberrant DNA methylation in CLL reported hypomethylation of two genes, *ornithine decarboxylase* (*ODC*) and *erb-A1*, when compared to leukocytes from healthy donors. Similarly, loss of DNA methylation in the *TCL1* promoter region was linked to elevated *TCL1* expression levels in CLL. Interestingly, *TCL1* overexpression was shown to contribute to CLL leukemogenesis as a result of Eμ-*TCL1* transgene expression; a finding which clearly demonstrated a causative role for *TCL1* promoter DNA hypomethylation in CLL pathogenesis. Likewise, the expression of many miRNAs in CLL was linked to aberrant epigenetic programming in their promoter regions. A study by Baer et al. reported 128 miRNAs that demonstrated DNA methylation changes in their promoters, a majority of which (e.g., miR-21, miR-34a, miR-29a, miR-9-2, miR-708, or miR-15) demonstrated loss of DNA methylation and consequently increased expression levels in CLL [34].

With the exception of a few reports on DNA hypomethylation in CLL, most studies focused on the understanding of promoter DNA hypermethylation. For example, hypermethylation of the promoter of the gene encoding the catalytic subunit of telomerase, *human telomerase reverse transcriptase* (*hTERT*), correlated with lower telomerase activity, longer telomeres, and superior overall survival as compared to the lowly methylated group of patients. Promoter hypermethylation of the transcription factor gene *TWIST2*, a negative regulator of p53, has been identified as a frequent event in *IGHV* mutated, good prognosis CLL cases (M-CLL) and correlated with lower *TWIST2* mRNA expression levels. Similarly, DNA hypermethylation of a single CpG within the *ZAP70* promoter has been linked to loss of *ZAP70* expression and a favorable clinical outcome. This finding paved the way toward the use of ZAP70 as a prognostic marker in CLL, helping to discriminate between poor prognostic *IGHV* unmutated (U-CLL) and favorable prognostic *IGHV* mutated (M-CLL) CLL subtypes. U-CLL patients show about sixfold higher *ZAP70* expression levels than M-CLL patients. *ZAP70* expression was shown to correctly predict *IGHV* mutation status in ~93% of CLL patients [34].

Methylomics identify CLL subgroups of different cellular origins

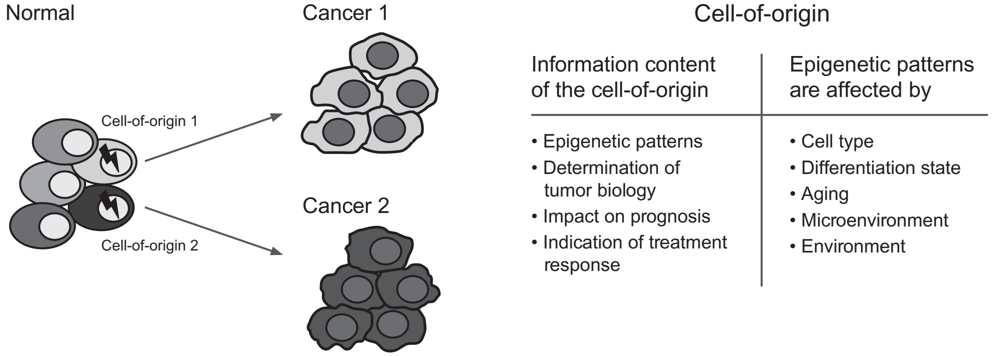
The advent of methylation arrays and next-generation sequencing technologies allowed genome-wide DNA methylation analysis at single CpG dinucleotide resolution. This led to the observation that CLL is characterized by widespread loss of DNA methylation, primarily targeting gene bodies, enhancer regions, and transcription factor binding sites. At the gene level, this DNA hypomethylation affects genes involved in NF-kB, B-cell receptor, and calcium-activated signaling. The few hypermethylated CpGs are enriched for 5′ regulatory regions, CpG islands, 5′ regions of introns, and Polycomb-repressed chromatin [35, 36].

Recent studies demonstrated that normal B-cell differentiation involves broad epigenetic programming that is mainly characterized by DNA hypomethylation. Based on this and the observation that similar patterns of methylation dynamics are detected in CLL, a novel epigenetic classification of CLL has been proposed based on their similarity to normally differentiating B-cell subpopulations. It is now accepted that CLL derives from a continuum of B-cell differentiation stages, rather than from one particular B-cell subtype [36–38]. Looking at the degree of methylation programming, all CLL samples closely resemble memory B cells, reaching around ~70%–100% of their developmental program. The overall level of achieved developmental programming was used to classify CLL into three groups, namely low-programmed CLL (LP-CLL) with poor prognosis, intermediate-programmed CLL (IP-CLL) with intermediate prognosis, and high-programmed CLL (HP-CLL) with good prognosis. LP-CLL exhibits the most immature methylation patterns, while in contrast, HP-CLL displays the most complete epigenetic programming resembling class-switched memory B cells. The CLL epitypes correlate well with clinically relevant parameters such as patient outcome, Binet stage, *SF3B1* mutation status, and *CD38* expression. This resulted in the development of DNA methylation classifiers using only a few CpGs that enable an accurate classification of CLL samples into one of the three epitypes [36, 37].

Identification of disease-specific DNA methylation patterns in CLL

The massive epigenetic programming during normal B-cell differentiation together with the observation that CLL can arise from different stages of normal B-cell differentiation urged the redefinition of how aberrant DNA methylation patterns are identified. It became clear that a simple comparison of CLL samples with normal CD19$^+$ B cells, which are mostly a mixture of naïve and mature memory B cells, will predominantly lead to the identification of DNA methylation changes that are in fact related to normal B-cell differentiation [36, 38]. In line with this, it has been demonstrated that many DNA methylation changes, which have been reported in CLL, are changes occurring during normal B-cell development and hence are not involved in CLL pathogenesis. The first attempt to identify cancer-specific DNA methylation changes has simply subtracted normal B-cell differentiation-related DNA methylation changes from CLL DNA methylomes. This identified aberrant DNA methylation programming to strongly affect NFAT and EGR binding motifs, and it implicated EGR2 abnormal activity as an important player in LP-CLL [38].

These findings nicely illustrate that the DNA methylome of any given CLL, and likely of any other tumor, results from a combination of DNA methylation patterns acquired during normal differentiation programming and cancer-specific events. The identification of cancer-specific DNA methylation patterns might lead to novel insights into disease biology but requires careful consideration and usage of normal reference DNA methylomes (Fig. 3).



Information content of the cell-of-origin	Epigenetic patterns are affected by
• Epigenetic patterns • Determination of tumor biology • Impact on prognosis • Indication of treatment response	• Cell type • Differentiation state • Aging • Microenvironment • Environment

Fig. 3 The cell of origin represents the cellular origin of the cancer-initiating cell, which is defined by its individual epigenetic state that is, amongst others, essentially affected by its individual differentiation state. Consequently, different kinds of differentiation states can give rise to different subgroups of cancers that differ significantly in tumor biology.

Epigenomics for disease classification and as a diagnostic tool

Large-scale epidemiologic studies have indicated that the worldwide cancer burden and number of cancer-related deaths are constantly increasing. This requires physicians and scientists to develop screening programs, which aim at detecting cancers at an early stage when patients have the highest chances to benefit from available treatment strategies. The discovery of cancer stage- and tumor type-specific biomarkers opened a new field of research with the promise of noninvasive cancer screening and accurate disease classification systems.

Epigenetic biomarkers

Cancer biomarkers can be classified in three main categories: prognostic biomarkers, which allow prediction of the disease course and survival chances; predictive biomarkers to assess, if a patient benefits from a certain treatment; and pharmacodynamic biomarkers, which are used in the clinics to guide personalized drug dosing and response assessment [39]. Several different biomarkers have been proposed, which depend on the underlying tumor entity, methodology as well as type and availability of patient material. At the very basic level, peripheral blood counts or other routine clinical laboratory parameters can be used to classify different leukemia entities or to approximate the risk of relapse and survival. More sophisticated methods require flow cytometry, cytochemistry, proteomic, and phosphoproteomic profiling by mass spectrometry, or the assessment of gene expression patterns and mutations by qRT-PCR, Sanger sequencing, or next-generation sequencing.

DNA methylation patterns, histone marks, and even higher-order chromatin structures have emerged as epigenetic cancer biomarkers during the last years. Especially DNA methylation qualifies as a cancer biomarker due to its inherent stability and the robustness of available laboratory assays. Feinberg and Vogelstein was reported in 1983 that cancers show global changes in their DNA methylation patterns. Genome-wide hypomethylation together with focal hypermethylation of tumor suppressor genes has been described as an epigenetic mechanism of oncogenesis [40, 41]. For instance, hypermethylation of the *MGMT* promoter has been frequently observed in human malignancies such as colorectal cancer, glioma, or non-small cell lung cancer. *MGMT* encodes the DNA repair enzyme O^6-methylguanine-DNA methyltransferase, which catalyzes the removal of alkyl-groups from the O^6 position of guanines and thereby protects cells from G-to-A transitions. The silencing of this gene by promoter hypermethylation is associated with increased sensitivity to alkylating drugs and an increased frequency of G-to-A mutations in oncogenes like *KRAS*. Clinical trials in glioma patients have already included methylation-specific PCR assays to correlate the promoter methylation status of *MGMT* with the response to temozolomide treatment together with radiotherapy.

An essential step in the establishment of an epigenetic biomarker is the development of a robust laboratory assay that can be used to measure incoming patient specimens. Such an assay should be applicable in clinical routine use, hence, ideally not requiring invasive procedures, and generating highly reproducible results with a minimum time, effort, and money spent. Besides DNA methylation arrays that analyze simultaneously thousands of CpGs, targeted epigenome profiling assays such as amplicon bisulfite sequencing or Epi-TYPER MassARRAY allow researchers to simultaneously and quantitatively determine the DNA methylation status of dozens of regions in the genome, which can already be enough information for robust classifications as discussed above. These methods have been proven to produce highly comparable results between different laboratories and allow the assessment of DNA methylation patterns of candidate loci at low cost [42].

The type of test material to use as well as the timing of testing need to be well defined. It is also important to consider potential confounding effects such as age, sex, ethnicity, interindividual genetic differences, environmental factors, or additional treatments. In case the biomarker seems to be affected by confounding factors, one should consider statistical adjustments.

DNA methylation cancer biomarkers can be identified in a variety of body fluids like sputum for lung cancer, urine for prostate cancer, or stool for colorectal cancer. For example, one study detected hypermethylation of p16 or MGMT in the sputum up to 3 years before the diagnosis of squamous cell lung carcinoma. More recently, cell-free DNA has been used for cancer diagnosis and classification, allowing the establishment of a DNA methylation biomarker based on minimally invasive procedures [43].

DNA methylation biomarkers for cancer classification and risk assessment

Currently, molecular disease classification in cancer largely depends on the knowledge of mutational patterns. Targeted sequencing panels and cytogenetic analysis allow physicians to quickly assess the frequency of common driver mutations and to assign personalized treatments to patients. However, several recent studies have established DNA methylation-based disease-classification systems for a variety of cancer entities, including malignancies of the central nervous system, sarcoma, and different forms of leukemia. In the case of CLL, hypomethylation of a single CpG dinucleotide in the regulatory region of *ZAP70* correlated with poor prognosis, indicating that a minimal methylation signature or even a single CpG dinucleotide might serve as a prognostic biomarker [44]. Genome-wide analysis of DNA methylation patterns in CLL has later on revealed three biologically and clinically distinct subgroups that correlated with disease prognosis. Based on this finding, a DNA methylation classifier has been developed, which used only five CpGs to accurately classify patients into one of the three CLL subgroups, thereby supporting the power of epigenetic biomarkers in deconvoluting the biological heterogeneity of CLL [37]. Similarly, a close correlation between epigenetic subgroups, prognosis, and disease phenotypes was also shown for other types of leukemia. An international consortium investigating DNA methylomes in patients with juvenile myelomonocytic leukemia (JMML) identified three DNA methylation signatures, which defined biologically distinct JMML subgroups that correlate with event-free survival after allogeneic stem cell transplantation. Using machine learning, a DNA methylation classifier has been developed, which allows prospective risk assessment and evaluation of personalized treatment strategies for JMML patients in the clinical setting [45, 46]. More globally, a DNA methylation classifier for the diagnosis and classification of central nervous system tumors has been developed, which classifies patients into 91 distinct cancer entities and subgroups [47]. Together, these studies underline the power of DNA methylation analysis in identifying different cancer entities based on their tissue- and cell type-specific DNA methylation signatures to provide personalized risk assessment in clinical practice.

Computational models for cancer classification

Modern oncology aims at developing personalized treatment strategies for cancer patients by incorporating genetic and epigenetic data into diagnostic routine workup. This requires advanced computational algorithms and analysis pipelines to enable fast and precise assessment of epigenetic alterations, mutations, and genomic rearrangements and to provide meaningful classifiers for molecular and prognostic subgroups.

The power of class prediction algorithms for cancer classification has already been demonstrated by pioneering studies using gene expression arrays to subclassify leukemia patients [48, 49]. The underlying models were able to accurately distinguish ALL from AML or mixed-lineage leukemia (MLL), respectively. Machine learning methods like random forest classifiers or support vector machines have been used since then to prospectively classify cancer patients in disease subgroups and to simultaneously translate the complexity of biomedical data into a human-readable classifier system.

Simple models based on regression or decision trees have been applied successfully for disease classification using DNA methylation array data. Nevertheless, these methods cannot keep pace with the complexity of the data generated. The development of deep learning algorithms solves this problem. Neural networks allow researchers to train models on large data sets without the need to extract relevant features beforehand. A deep neural network consists of several layers, which are specifically designed to process different types of input data. The specific type and number of layers, their interconnection and the overall architecture of a neural network depend on the data structure [50]. Deep neural networks have been used for cancer classification, e.g., by scanning electronic health care reports (EHR), for the analysis of X-ray scans or pathology slides, and for the diagnosis of skin cancer from images. An important step in assessing the power of machine learning algorithms is to compare them to classical disease classification systems. The performance and accuracy of such algorithms are often measured by the "area under the ROC curve" (AUC). Neural networks often achieve an AUC of more than 0.9 in classifying images of skin lesions of pathology slides. However, the added value of a computational method to classic disease classification highly depends on the improvement of clinical outcomes. Many current studies lack the incorporation of models in clinical routine workflows, thereby neglecting the final measure, which assesses the real power of machine learning models [51].

Although the power of epigenetic biomarkers has been demonstrated in several studies, to date, their implementation in routine clinical work-up or in cancer screening programs is still missing. By using state-of-the-art machine learning algorithms and integrative approaches, clinicians will soon be able to accurately classify patients in disease subgroups and assign personalized treatment strategies based on the genetic and epigenetic abnormalities of cancers.

Epigenetic therapies in hematopoietic neoplasms

As outlined before, epigenetic mechanisms play a central role in the development and progression of cancer and therefore constitute attractive targets for therapeutic approaches. The reversible nature of epigenetic patterns combines advantages and disadvantages in terms of the establishment of therapeutic targets. On one hand, epigenetic features are theoretically easier to modulate than genetic mutations. On the other hand, the stability of such manipulations is presumably less robust. Nevertheless, the field of epigenetic therapies is growing rapidly. This is reflected in a large number of compounds that target all kinds of epigenetic players including writers, readers, and erasers [52]. The Human Epigenetic Drug Database (HEDD) catalogs seven categories of epigenetic drugs, namely DNA methyltransferase inhibitors (DNMTi), histone acetyltransferase inhibitors (HATi), histone demethylase inhibitors (HDMi), histone deacetylase inhibitors (HDACi), histone methyltransferase inhibitors (HMTi), inhibitors of protein binding to acetylated histones (PAHi), and inhibitors of protein binding to methylated histones (PMHi) [53]. Of note, all of these groups, except for PMHi, contain compounds that are being tested in several types of leukemia.

Inhibitors of epigenetic key players

In hematopoietic neoplasms, epigenetic regulators are frequently affected by mutations. This coincides with the fact that several drugs targeting epigenetic regulators are approved for the treatment of hematopoietic neoplasms and many more are being tested in preclinical and clinical trials. In October 2016, seven epigenetic drugs were approved for the treatment of malignant diseases, all of which have been approved for the treatment of hematopoietic neoplasms [54]. Recently, a novel class of epigenetic drugs, so-called mutant IDH inhibitors (mIDHi), has been developed. Two mIDHi, ivosidenib and enasidenib, have been approved, and several other mIDHi are being tested in preclinical and clinical trials. Overall, more than 100 compounds targeting epigenetic players are actively being tested in more than 1400 clinical trials.

All currently approved epigenetic drugs are inhibitors of DNA methyltransferases, inhibitors of histone deacetylases, or inhibitors of mutant forms of IDH1/2. DNMTi is cytosine analogs, which are incorporated into nucleic acids (RNA and/or DNA) and covalently bind methyltransferases. Both azacitidine and decitabine are DNMTi that are approved for AML, chronic myeloid leukemia (CML), and myelodysplastic syndromes (MDS) [54]. High doses of DNMTi are cytotoxic, probably by inhibiting DNA synthesis and hence blocking cell cycle progression. On the contrary, low doses of DNMTi efficiently trap DNMT enzymes and thereby promote their proteasomal degradation. This leads to a gradual decrease in DNA methylation that largely depends on the number of cell cycles. Drug-induced hypomethylation recovers quickly after drug discontinuation, and, in consequence, therapeutic response to DNMTi is relatively slow and reversible. Current efforts aim at developing less toxic DNMTi, which include antisense oligonucleotides, non-nucleoside DNMTi, and refined nucleoside analogs [55].

The second group of approved epigenetic drugs consists of inhibitors of histone deacetylases (HDAC). HDACi can be subdivided into isoenzyme-selective inhibitors that target-specific HDAC enzymes and pan-inhibitors that act on some but not all HDACs [55]. HDACs are overexpressed in several cancers and are proposed to have oncogenic potential. Therefore, the set of currently five approved HDAC inhibitors will likely be extended soon. Paradoxically, HDACi does not only repress but also activate transcription, which might to some degree be caused by the ability of HDACs to deacetylate not only histone proteins but also other proteins such as regulators of transcription and apoptosis. In consequence, the therapeutic mechanism of HDACi is hard to dissect. HDACi have been approved in cutaneous and peripheral T-cell lymphomas and multiple myeloma (MM), while they show modest activity in myelodysplastic syndromes (MDS), AML, and Hodgkin's lymphoma (HL) [56].

Inhibitors of mutant IDH enzymes have recently been developed and constitute a novel class of epigenetic drugs that exert indirect epigenetic effects. Two compounds of this class, ivosidenib, and enasidenib, have recently been approved for the treatment of relapsed or refractory AML with *IDH1* or *IDH2* mutations, respectively [56–58]. Clinical development of mIDHi is ongoing for several other cancer entities such as myelodysplastic syndromes, cholangiocarcinoma, glioma, and other solid tumors. Mutant forms of IDH1/2 are promising therapeutic targets as they are frequently mutated in different cancer entities and they are linking cellular metabolism to epigenetic dysregulation. This pleiotropy presumably explains the clinical potency of mIDHi. Remarkable is the specificity of those drugs for the mutant forms of these enzymes, while the wild-type proteins remain unaffected. The development of drugs targeting only the cancerous forms of key oncological factors opens a therapeutic window for otherwise likely toxic treatments, and refinement of target selectivity might further reduce the frequency and severity of treatment-related side effects.

Several other classes of epigenetic drugs are currently being evaluated in preclinical and clinical studies. For example, BET inhibitors interrupt protein-protein interactions between BET proteins and acetylated histones. BET inhibitors show clinical activity in a rare cancer entity called NUT midline carcinoma as well as in some subtypes of AML, multiple myeloma (MM), and acute lymphoblastic leukemia (ALL) [55, 56, 59]. One potential mechanism for antitumor activity of BET inhibitors consists of their ability to downregulate MYC expression.

The enzymatic component of the PRC2 complex, EZH2, mediates crosstalk in epigenetic regulation and is frequently altered in several cancer types including different types of leukemia. First-generation inhibitors lacked specificity for EZH2 and also inhibited other HMTs. The important role of EZH2 in diseases such as B-cell lymphomas, however, motivated intensive investigations on EZH2 as a pharmacological target. Currently, more specific drugs are being tested in preclinical and clinical studies for the treatment of non-Hodgkin lymphoma, ALL, MM, mesothelioma, and synovial sarcoma [59].

Inhibitors of the H3K79 methyltransferase DOT1L are another class of epigenetic drugs, which are evaluated in mixed-lineage leukemia (MLL) and might also be potent in DNMT3A- and NPM1-mutated AML [56, 59]. A potency for AML treatment has also been suggested for inhibitors targeting the H3K36 demethylase LSD1 [56].

Mechanisms of epigenetic therapies and possible rationales for drug combinations

Classically, epigenetic therapies like DNMTi and HDACi were believed to act via reactivation of tumor suppressor genes (TSG), but it became clear that this is more than an oversimplification of the real situation. Consequently, the lack of understanding of the pharmacological mechanism of action of DNMTi and HDACi is the strongest point of critique against epigenetic therapies. However, recent mechanistic insights into the molecular effects of DNMTi and HDACi suggest pleiotropic effects in cancer cells and reveal surprisingly specific effects on transcriptional regulation [60]. The global effects on chromatin dynamics and consequently transcriptional regulation appear to affect crucial cell biological mechanisms such as apoptosis and cell cycle regulation. In consequence, this can lead to immune-stimulatory effects [61], which opens promising novel therapeutic strategies. DNMTi and HDACi therapy result in the upregulation of antigen generation, processing, and presentation, which consequently upregulates interferon-mediated inflammation. It has been proposed that such effects create stable targets that remain to be expressed beyond the time of acute treatment [60, 62]. In addition, reduced autophagy and angiogenesis have been described as consequences of HDACi therapy. Strikingly, the combination of DNMTi and HDACi triggers a massive activation of specific repetitive elements, which generate thousands of cryptic transcripts. A subset of these transcripts is supposed to give rise to novel antigens that might function as targets for a combined immune therapy [60]. Currently, both HDACi (Vorinostat) and DNMTi (Decitabine) are clinically tested in combination with immune checkpoint inhibitors. This and other clinical trials combining DNMTi and HDACi with immunotherapeutic approaches will establish whether such combinations really translate into clinically meaningful effects [55, 60, 62].

Furthermore, epigenetic therapies can be combined with hormonal or cytotoxic drugs as well as with radiotherapy. It is exactly this what comprises a very promising clinical development in oncology. Epigenetic therapies can be combined with further epigenetic drugs and, theoretically, all other kinds of available therapies. Until 2016, more than half of the completed trials were designed on combinations. This number raised to two-third of all ongoing and planned clinical trials evaluating combinatorial treatments of mostly two to three and up to seven drugs at once [54]. The current interpretation of the efficacy of such approaches is that epigenetic treatments trigger a priming effect that sensitizes cancer cells to further treatments. This affects immunogenicity, cell growth, DNA damage, and apoptosis.

In the light of emergent combinatorial therapies, the disadvantage that epigenetic manipulations are transient might soon transform into an advantage. The epigenetic sensitization of cancer cells in addition to a long list of different therapeutic options might strongly increase both the specificity and range of applications at the same time. The concept of cancer stem cells acting as both initiators and critical promoters of the malignant disease put additional emphasis on epigenetics. The cell of origin can be defined by epigenetic profiling and provides information on disease-specific epigenetic aberrations. Together this knowledge might reveal novel therapeutic targets to attack highly resistant cell populations such as leukemia stem cells.

In conclusion, the field of epigenetic therapy is growing rapidly, and the examples mentioned here recapitulate the ongoing widespread activities with respect to the development of novel compounds targeting epigenetic regulators. It can be anticipated that we will victim a plethora of novel agents interfering with epigenetic regulation. The clinical and scientific challenge will be to find the appropriate drug combinations and dosing schedules to inhibit the therapeutic targets in order to optimally exploit their therapeutic potential.

Acknowledgment

We would like to thank Magdalena Hartmann for supporting the graphic design of the figures.

References

[1] Haas S, Trumpp A, Milsom MD. Causes and consequences of hematopoietic stem cell heterogeneity. Cell Stem Cell 2018;22(5):627–38.
[2] Stricker SH, Koferle A, Beck S. From profiles to function in epigenomics. Nat Rev Genet 2017; 18(1):51–66.

[3] Laurenti E, Gottgens B. From haematopoietic stem cells to complex differentiation landscapes. Nature 2018;553(7689):418–26.

[4] Broske AM, Vockentanz L, Kharazi S, Huska MR, Mancini E, Scheller M, et al. DNA methylation protects hematopoietic stem cell multipotency from myeloerythroid restriction. Nat Genet 2009; 41(11):1207–15.

[5] Bock C, Beerman I, Lien WH, Smith ZD, Gu H, Boyle P, et al. DNA methylation dynamics during in vivo differentiation of blood and skin stem cells. Mol Cell 2012;47(4):633–47.

[6] Cabezas-Wallscheid N, Klimmeck D, Hansson J, Lipka DB, Reyes A, Wang Q, et al. Identification of regulatory networks in HSCs and their immediate progeny via integrated proteome, transcriptome, and DNA methylome analysis. Cell Stem Cell 2014;15(4):507–22.

[7] Farlik M, Halbritter F, Muller F, Choudry FA, Ebert P, Klughammer J, et al. DNA methylation dynamics of human hematopoietic stem cell differentiation. Cell Stem Cell 2016;19(6):808–22.

[8] Buenrostro JD, Corces MR, Lareau CA, Wu B, Schep AN, Aryee MJ, et al. Integrated single-cell analysis maps the continuous regulatory landscape of human hematopoietic differentiation. Cell 2018; 173(6):1535–1548.e16.

[9] Lara-Astiaso D, Weiner A, Lorenzo-Vivas E, Zaretsky I, Jaitin DA, David E, et al. Immunogenetics. Chromatin state dynamics during blood formation. Science (New York, NY) 2014;345(6199):943–9.

[10] Lee-Six H, Obro NF, Shepherd MS, Grossmann S, Dawson K, Belmonte M, et al. Population dynamics of normal human blood inferred from somatic mutations. Nature 2018;561(7724):473–8.

[11] Welch JS, Ley TJ, Link DC, Miller CA, Larson DE, Koboldt DC, et al. The origin and evolution of mutations in acute myeloid leukemia. Cell 2012;150(2):264–78.

[12] Langstein J, Milsom MD, Lipka DB. Impact of DNA methylation programming on normal and preleukemic hematopoiesis. Semin Cancer Biol 2018;51:89–100.

[13] Desai P, Mencia-Trinchant N, Savenkov O, Simon MS, Cheang G, Lee S, et al. Somatic mutations precede acute myeloid leukemia years before diagnosis. Nat Med 2018;24(7):1015–23.

[14] Rossi DJ, Bryder D, Zahn JM, Ahlenius H, Sonu R, Wagers AJ, et al. Cell intrinsic alterations underlie hematopoietic stem cell aging. Proc Natl Acad Sci U S A 2005;102(26):9194–9.

[15] Ley TJ, Ding L, Walter MJ, McLellan MD, Lamprecht T, Larson DE, et al. DNMT3A mutations in acute myeloid leukemia. N Engl J Med 2010;363(25):2424–33.

[16] Cole CB, Russler-Germain DA, Ketkar S, Verdoni AM, Smith AM, Bangert CV, et al. Haploinsufficiency for DNA methyltransferase 3A predisposes hematopoietic cells to myeloid malignancies. J Clin Invest 2017;127:3657–74.

[17] Moran-Crusio K, Reavie L, Shih A, Abdel-Wahab O, Ndiaye-Lobry D, Lobry C, et al. Tet2 loss leads to increased hematopoietic stem cell self-renewal and myeloid transformation. Cancer Cell 2011; 20(1):11–24.

[18] Corces-Zimmerman MR, Hong WJ, Weissman IL, Medeiros BC, Majeti R. Preleukemic mutations in human acute myeloid leukemia affect epigenetic regulators and persist in remission. Proc Natl Acad Sci U S A 2014;111(7):2548–53.

[19] Xu W, Yang H, Liu Y, Yang Y, Wang P, Kim SH, et al. Oncometabolite 2-hydroxyglutarate is a competitive inhibitor of alpha-ketoglutarate-dependent dioxygenases. Cancer Cell 2011;19(1):17–30.

[20] Abdel-Wahab O, Adli M, LaFave LM, Gao J, Hricik T, Shih AH, et al. ASXL1 mutations promote myeloid transformation through loss of PRC2-mediated gene repression. Cancer Cell 2012; 22(2):180–93.

[21] Dohner H, Estey E, Grimwade D, Amadori S, Appelbaum FR, Buchner T, et al. Diagnosis and management of AML in adults: 2017 ELN recommendations from an international expert panel. Blood 2017;129(4):424–47.

[22] Youn BS, Mantel C, Broxmeyer HE. Chemokines, chemokine receptors and hematopoiesis. Immunol Rev 2000;177:150–74.

[23] Figueroa ME, Lugthart S, Li Y, Erpelinck-Verschueren C, Deng X, Christos PJ, et al. DNA methylation signatures identify biologically distinct subtypes in acute myeloid leukemia. Cancer Cell 2010; 17(1):13–27.

[24] Martens JH, Stunnenberg HG. The molecular signature of oncofusion proteins in acute myeloid leukemia. FEBS Lett 2010;584(12):2662–9.

[25] Sanjuan-Pla A, Bueno C, Prieto C, Acha P, Stam RW, Marschalek R, et al. Revisiting the biology of infant t(4;11)/MLL-AF4+ B-cell acute lymphoblastic leukemia. Blood 2015;126(25):2676–85.

[26] Koh CP, Wang CQ, Ng CE, Ito Y, Araki M, Tergaonkar V, et al. RUNX1 meets MLL: epigenetic regulation of hematopoiesis by two leukemia genes. Leukemia 2013;27(9):1793–802.

[27] Jerez A, Sugimoto Y, Makishima H, Verma A, Jankowska AM, Przychodzen B, et al. Loss of heterozygosity in 7q myeloid disorders: clinical associations and genomic pathogenesis. Blood 2012; 119(25):6109–17.

[28] Mrozek K, Eisfeld AK, Kohlschmidt J, Carroll AJ, Walker CJ, Nicolet D, et al. Complex karyotype in de novo acute myeloid leukemia: typical and atypical subtypes differ molecularly and clinically. Leukemia 2019;33(7):1620–34.

[29] Belver L, Ferrando A. The genetics and mechanisms of T cell acute lymphoblastic leukaemia. Nat Rev Cancer 2016;16(8):494–507.

[30] Girardi T, Vicente C, Cools J, De Keersmaecker K. The genetics and molecular biology of T-ALL. Blood 2017;129(9):1113–23.

[31] Peirs S, Van der Meulen J, Van de Walle I, Taghon T, Speleman F, Poppe B, et al. Epigenetics in T-cell acute lymphoblastic leukemia. Immunol Rev 2015;263(1):50–67.

[32] Navarro J-M, Touzart A, Pradel LC, Loosveld M, Koubi M, Fenouil R, et al. Site- and allele-specific polycomb dysregulation in T-cell leukaemia. Nat Commun 2015;6(1).

[33] Mansour MR, Abraham BJ, Anders L, Berezovskaya A, Gutierrez A, Durbin AD, et al. Oncogene regulation. An oncogenic super-enhancer formed through somatic mutation of a noncoding intergenic element. Science (New York, NY) 2014;346(6215):1373–7.

[34] Mansouri L, Wierzbinska JA, Plass C, Rosenquist R. Epigenetic deregulation in chronic lymphocytic leukemia: clinical and biological impact. Semin Cancer Biol 2018;51:1–11.

[35] Kulis M, Merkel A, Heath S, Queiros AC, Schuyler RP, Castellano G, et al. Whole-genome fingerprint of the DNA methylome during human B cell differentiation. Nat Genet 2015;47(7):746–56.

[36] Kulis M, Heath S, Bibikova M, Queiros AC, Navarro A, Clot G, et al. Epigenomic analysis detects widespread gene-body DNA hypomethylation in chronic lymphocytic leukemia. Nat Genet 2012;44(11):1236–42.

[37] Queiros AC, Villamor N, Clot G, Martinez-Trillos A, Kulis M, Navarro A, et al. A B-cell epigenetic signature defines three biologic subgroups of chronic lymphocytic leukemia with clinical impact. Leukemia 2015;29(3):598–605.

[38] Oakes CC, Seifert M, Assenov Y, Gu L, Przekopowitz M, Ruppert AS, et al. DNA methylation dynamics during B cell maturation underlie a continuum of disease phenotypes in chronic lymphocytic leukemia. Nat Genet 2016;48(3):253–64.

[39] Sawyers CL. The cancer biomarker problem. Nature 2008;452(7187):548–52.

[40] Feinberg AP, Vogelstein B. Hypomethylation distinguishes genes of some human cancers from their normal counterparts. Nature 1983;301(5895):89–92.

[41] Greger V, Passarge E, Hopping W, Messmer E, Horsthemke B. Epigenetic changes may contribute to the formation and spontaneous regression of retinoblastoma. Hum Genet 1989;83(2):155–8.

[42] BLUEPRINT Consortium. Quantitative comparison of DNA methylation assays for biomarker development and clinical applications. Nat Biotechnol 2016;34(7):726–37.

[43] Shen SY, Singhania R, Fehringer G, Chakravarthy A, Roehrl MHA, Chadwick D, et al. Sensitive tumour detection and classification using plasma cell-free DNA methylomes. Nature 2018;563(7732): 579–83.

[44] Claus R, Lucas DM, Ruppert AS, Williams KE, Weng D, Patterson K, et al. Validation of ZAP-70 methylation and its relative significance in predicting outcome in chronic lymphocytic leukemia. Blood 2014;124(1):42–8.

[45] Lipka DB, Toth R, Yang J, Wiesenfarth M, Nöllke P, Fischer A, Brocks D, Gu Z, Park J, Strahm B, Wlodarski M, Yoshimi A, Claus R, Lübbert M, Busch H, Boerries M, Hartmann M, Schönung M, Kilik U, Langstein J, Wierzbinska J, Pabst C, Garg S, Català A, De Moerloose B, Dworzak M, Hasle H, Locatelli F, Masetti R, Schmugge M, Smith O, Stary J, Ussowicz M, van den Heuvel-Eibrink M, Assenov Y, Schlesner M, Niemeyer C, Flotho C, Plass C. RAS-pathway mutation patterns define epigenetic subclasses in juvenile myelomonocytic leukemia. Nat Commun 2017;8:2126.

[46] Schoenung M, Stieglitz E, Muramatsu H, Murakami N, Hartmann M, Wiesenfarth M, et al. DNA methylation subgroups in juvenile myelomonocytic leukemia: an international collaborative analysis and development of a common diagnostic platform. Blood 2018;132.

[47] Capper D, Jones DTW, Sill M, Hovestadt V, Schrimpf D, Sturm D, et al. DNA methylation-based classification of central nervous system tumours. Nature 2018;555(7697):469–74.

[48] Armstrong SA, Staunton JE, Silverman LB, Pieters R, den Boer ML, Minden MD, et al. MLL translocations specify a distinct gene expression profile that distinguishes a unique leukemia. Nat Genet 2002;30(1):41–7.

[49] Golub TR, Slonim DK, Tamayo P, Huard C, Gaasenbeek M, Mesirov JP, et al. Molecular classification of cancer: class discovery and class prediction by gene expression monitoring. Science (New York, NY) 1999;286(5439):531–7.

[50] Yu KH, Beam AL, Kohane IS. Artificial intelligence in healthcare. Nat Biomed Eng 2018;2(10): 719–31.

[51] Topol EJ. High-performance medicine: the convergence of human and artificial intelligence. Nat Med 2019;25(1):44–56.

[52] Prachayasittikul V, Prathipati P, Pratiwi R, Phanus-Umporn C, Malik AA, Schaduangrat N, et al. Exploring the epigenetic drug discovery landscape. Expert Opin Drug Discovery 2017;12(4):345–62.

[53] Qi Y, Wang D, Wang D, Jin T, Yang L, Wu H, et al. HEDD: the human epigenetic drug database. Database 2016;2016.

[54] Meighan-Mantha R. Epigenetic drugs in oncology: current clinical landscape and emerging trends. Cancer 2017;1–16.

[55] Daskalakis M, Brocks D, Sheng YH, Islam MS, Ressnerova A, Assenov Y, et al. Reactivation of endogenous retroviral elements via treatment with DNMT- and HDAC-inhibitors. Cell Cycle 2018; 17(7):811–22.

[56] Goyama S, Kitamura T. Epigenetics in normal and malignant hematopoiesis: An overview and update 2017. Cancer Sci 2017;108(4):553–62.

[57] Wouters BJ, Delwel R. Epigenetics and approaches to targeted epigenetic therapy in acute myeloid leukemia. Blood 2016;127(1):42–52.

[58] Newcombe AA, Gibson BES, Keeshan K. Harnessing the potential of epigenetic therapies for childhood acute myeloid leukemia. Exp Hematol 2018;63:1–11.

[59] Popovic R, Shah MY, Licht JD. Epigenetic therapy of hematological malignancies: where are we now? Ther Adv Hematol 2013;4(2):81–91.

[60] Brocks D, Schmidt CR, Daskalakis M, Jang HS, Shah NM, Li D, et al. DNMT and HDAC inhibitors induce cryptic transcription start sites encoded in long terminal repeats. Nat Genet 2017;49(7): 1052–60.

[61] Bewersdorf JP, Shallis R, Stahl M, Zeidan AM. Epigenetic therapy combinations in acute myeloid leukemia: what are the options? Ther Adv Hematol 2019;10. 2040620718816698.

[62] Chiappinelli KB, Strissel PL, Desrichard A, Li H, Henke C, Akman B, et al. Inhibiting DNA methylation causes an interferon response in cancer via dsRNA including endogenous retroviruses. Cell 2015;162(5):974–86.

This page intentionally left blank

CHAPTER 14

Impact of epigenetic modifiers on the immune system

Sajad Ahmad Bhat[a,b], **Shalini Kashipathi Sureshbabu**[a,b], **Chinna Susan Philip**[a,b], **Shubhada Chiplunkar**[a,b]
[a]Chiplunkar Laboratory, Advanced Centre for Treatment, Research and Education in Cancer (ACTREC), Tata Memorial Centre, Mumbai, India
[b]Homi Bhabha National Institute, Mumbai, India

Contents

Overview 315
Epigenetic modifiers 316
    HDAC inhibitors 317
    HAT inhibitors 317
    DNMT inhibitors 317
    BET inhibitors 317
    EZH2 inhibitors 318
    LSD1 inhibitors 319
    DOT1L inhibitors 319
Immunomodulatory effects of epigenetic modifiers 319
    Effect of HDACi on immune cells 321
    Effect of DNMTi on immune cells 324
Clinical significance of epigenetic modifiers 325
    HDACi therapy 325
    DNMTi therapy 341
    Other epigenetic modifiers in therapy 341
Combination therapy involving epigenetic modifiers 342
    Epigenetic modifiers and immunotherapy 342
    Epigenetic modifiers and chemotherapeutics 343
    Epigenetic modifiers and radiotherapy 344
    Combination therapy with HDACi and DNMTi 344
Conclusion 345
Acknowledgments 346
References 346

Overview

The genetic information within eukaryotic cells is stored as the nucleotide sequence of DNA which is further densely packed into chromatin in association with histone and nonhistone proteins. The physiological function of a cell depends on the accessibility of this genetic information [1, 2]. The more the DNA is tightly wound, the more the

Epigenetics of the Immune System
https://doi.org/10.1016/B978-0-12-817964-2.00014-9
© 2020 Elsevier Inc.
All rights reserved.

DNA is condensed, and gene expression is repressed. Modulation of gene transcription is a key factor in numerous biological processes [3]. Aberrant gene expression results in the onset of various pathological conditions [4]. Epigenetic factors which involve different epigenetic enzymes play an important role in producing posttranslational modifications on histone proteins as well as the methylation of the DNA sequence, resulting in alterations in the accessibility and chromatin structure [5]. These modifications regulate the diverse patterns of gene expression, thereby controlling various cellular processes, including gene and microRNA expression, embryogenesis, cellular differentiation, DNA-protein interactions, X-chromosome inactivation, suppression of transposable element mobility, and genomic imprinting [6]. Histones undergo posttranslational modifications such as methylation, acetylation, phosphorylation, ubiquitylation, and ADP ribosylation that are controlled by specific epigenetic enzymes which have a significant impact on the immune system and its responses, the details of which have been explained in the previous chapters. The modulation of these enzymes using their inhibitors also called epigenetic modifiers and their clinical significance are discussed in detail in the subsequent sections.

Aberrant epigenetic modifications such as DNA methylation and histone acetylation have been well described in many human diseases. These aberrant epigenetic modifications are due to deregulation or mutations of different epigenetic players. The reversible nature of these epigenetic modifications makes them good targets for therapeutic intervention in different human diseases. In the past few decades epigenetic-based therapies are being considered as a treatment modality in diseases like cancer. Targeting epigenetic players with specific inhibitors such as DNA methyltransferase (DNMT) inhibitors, histone acetyltransferase (HAT) inhibitors, and histone deacetylases (HDAC) inhibitors play a crucial role in modulating the expression of key genes and other biological processes. Targeting the different epigenetic enzymes with these modifiers modulates the immunogenicity and response of the infected cells. Besides the first clinically approved epigenetic modifiers, DNMT inhibitors and HDAC inhibitors, several new classes of epigenetic modifiers such as bromodomain and extra-terminal motif (BET) inhibitors, enhancer of zeste homolog 2 (EZH2) inhibitors, lysine-specific demethylase 1 (LSD1) inhibitors, disruptor of telomeric silencing 1-like (DOT1-L) and arginine *N*-methyltransferase 5 (PRMT5) inhibitors have been recently discovered and are in early clinical development.

Epigenetic modifiers

Because of the strong correlation between epigenetic posttranslational modifications and the development of different diseases, various inhibitors of the epigenetic enzymes have been developed and are being investigated in clinical trials for therapeutic gains. These inhibitors or epigenetic modifiers are expected to develop as novel strategies to combat life-threatening pathological conditions.

HDAC inhibitors

Histone deacetylase (HDAC) inhibitors basically act against the HDAC enzymes to modify the acetylation of histones in chromatin and augment gene transcription [7]. They alter gene expression directly or indirectly by posttranslational acetylation of histone and nonhistone proteins [8]. HDAC inhibitors block the Zn^{2+} binding domain on the HDACs, thereby affecting their catalytic activity [7]. Based on their structural differences, HDAC inhibitors have been classified into five different groups (I) hydroxamic acids (hydroxamates) which include trichostatin A (TSA) and suberoylanilide hydroxamic acid (SAHA); (II) short-chain fatty (aliphatic) acids which include valproic acid (VPA) and butyric acid; (III) benzamides such as entinostat and tacedinaline; (IV) cyclic tetrapeptides such as romidepsin; and (V) sirtuin (SIRT) inhibitors including the pan-inhibitor nicotinamide and specific SIRT1 and SIRT2 inhibitors such as sirtinol and cambinol, respectively [9].

HAT inhibitors

HAT inhibitors hinder the catalytic activity of HATs, i.e., they inhibit the transfer of acetyl group on histones. They have been classified as natural products (anacardic acid from cashew nuts and curcumin), bi-substrate inhibitors (cofactor acetyl CoA and lysine-containing substrate), and synthetic small molecules (γ-butyrolactone MB-3, isothiazolones, garcinol analogues) [10, 11]. Although these inhibitors show potential results against various diseases, the molecular basis of their inhibition still needs to be elucidated.

DNMT inhibitors

DNMT inhibitors act by reactivation of the genes aberrantly silenced via methylation by DNMTs and restoration of their normal function. These inhibitors covalently bind DNMTs by incorporating themselves into the growing DNA strand. DNMT inhibitors are classified into two categories—nucleoside and non-nucleoside inhibitors. Nucleoside inhibitors which include 5-azacitidine (5-Aza) and 5-aza-2′-deoxycytidine (decitabine) are usually derived from cytidine and require integration in the DNA to be active. Non-nucleoside inhibitors such as curcumin and procaine have variation in their chemical structure and in general bind directly to DNMTs [12].

BET inhibitors

Members of bromodomain and extra-terminal (BET) are a subfamily of bromodomain proteins (BRD) which include BRD2, BRD3, and BRD4. These proteins are epigenetic readers mainly associated with the acetylated chromatin and are responsible for transcriptional activation by recruiting different transcriptional factors through protein-protein interaction [13]. The abrupt fusion of BET proteins with different transcription factors BRD4-NUT fusion is responsible for rare human cancer NUT Midline carcinoma which is an aggressive subtype of squamous cell carcinoma with a median survival of only 6.7 months [14]. At present at least 10 BET inhibitors are under early clinical investigation. Recent research suggested that BET inhibitors have a good impact on different malignancies. Two well-known selective BET inhibitors include JQ1 and I-BET762. A recent study by Asangani et al. has shown that JQ1 inhibits the binding of androgen receptor to target gene by inhibiting the BRD4 and androgen receptor interaction in prostate cancer [15]. I-BET762 has been shown to inhibit the proliferative potential of pancreatic cancer cells [16]. Besides this antitumor activity, BET inhibitors also have immunomodulatory functions. BET inhibitor I-BET762 inhibits nitric oxide (NO) production and other proinflammatory cytokines [17]. A recent study by Zhu et al. has demonstrated that PD-L1 is a direct target of the BRD4 which regulated the expression of PD-L1 in ovarian cancer and associated immune cells [18]. Treatment with BET inhibitors suppresses the expression of PD-L1 in these cells and thus modulates the immune response in a favorable manner [18]. BET inhibitors decreased the tumor progression in ovarian cancer mouse model by increasing the production of proinflammatory cytokine interferon-γ (IFN-γ) and effector molecule granzyme B in tumor-infiltrating CD8+ T cells [19]. Although BET inhibitors have shown promising results in preclinical mouse models, clinical trials on human subjects are still ongoing.

EZH2 inhibitors

Histone methyltransferease Enhancer of zeste homolog 2 (EZH2) is the catalytic component of polycomb repressive complex 2 (PRC2) and catalyzes the methylation of lysine residue 27 of histone H3 (H3K27 methylation). Besides playing a role in different immunological functions, dysregulation of EZH2 is observed in different human malignancies such as breast cancer, ovarian cancer, prostate cancer, non-Hodgkin lymphoma (NHL), and T-cell acute lymphoblastic leukemia (T-ALL) [20]. Inhibition of the catalytic activity of EZH2 has emerged as a novel therapeutic approach to target these different human cancers. This pharmacologic role of EZH2 has led to the development of specific EZH2 inhibitors recently such as tazemetostat, GSK2816126, and CPI-1205. Due to the promising results in xenograft models, these inhibitors have moved tophase I and phase II clinical trials for the treatment of different tumors. Early clinical trial data from the EZH2 inhibitor tazemetostat has shown encouraging responses in lymphoma patients [21]. Combinatorial therapy of EZH2 inhibitors with HDAC inhibitor SAHA and PD-1 inhibitor atezolizumab have also shown promising results in pediatric patients with lymphoma or genetically defined solid tumors [22]. Further understanding of the outcome of combinatorial therapy of EZH2 with different conventional chemotherapeutic and other agents is necessary for the designing of new successful treatment strategies in different pathological conditions like cancer.

LSD1 inhibitors

Histone demethylases especially lysine-specific demethylases (LSDs) are transcription regulators that regulate the gene expression by demethylating histone H3 on lysine 4 and lysine 9 (H3K4 and H3K9) [23]. LSD1 is involved in regulating the expression of different proinflammatory cytokines and aberrant expression of these cytokines is associated with different diseases [23, 24]. Treatment involving LSD1 inhibition along with HDAC inhibition restores the antitumor immunity [25]. LSD1 is overexpressed in different types of tumors and is involved in the proliferation, epithelial-mesenchymal transition, stem cell biology, and malignant transformation of cells [26]. Thus inhibition of LSD1 by different inhibitors is becoming the attractive target for novel therapeutic treatments. High expression of LSD1 in ovarian cancers is responsible for high aggressive tumors due to downregulation of the genes involved in the immune and inflammatory responses [27]. New LSD1 inhibitors that are currently in early development may potentially be used as a treatment for cancer in future.

DOT1L inhibitors

Histone methyltransferase Disruptor of telomeric silencing 1-like (DOT1L) is involved in the mono-, di-, and trimethylation of the ε-amino group on H3K79 by means of *S*-(5′-adenosyl)-l-methionine (SAM) as the cosubstrate [28]. It has a significant role in DNA repair, transcriptional elongation, cell cycle regulation, cardiac function, erythropoiesis, and development of leukemia [29–33]. Dysregulation of DOT1L has been found to lead to the development of mixed lineage leukemia 1 (MLL1) rearranged leukemias [34]. Since its discovery, there have been various studies undertaken to find potential DOT1L inhibitors for therapeutic purposes. These inhibitors have been classified as SAH(*S*-adenosylhomocysteine)-like, benzimidazole/urea-containing, carbamate-containing, and mechanism-based DOT1L inhibitors [35]. These inhibitors are still emerging and further investigation needs to be carried out to exploit their potential as a targeted cancer therapeutic approach.

The catalytic activity of different epigenetic enzymes and their inhibitors (epigenetic modifiers) have a significant role in the regulation and function of the immune system. They can modulate the development, differentiation, and effector functions of the various immune cells thereby playing a major role in the immune responses of the body.

Immunomodulatory effects of epigenetic modifiers

Epigenetic modifiers have been found to play a significant role in the regulation of chromatin remodeling and gene expression by modulating the functions of numerous transcription factors and nonhistone proteins [36]. The addition of functional groups to histone tails, firstly, provides recruitment signals for the nonhistone proteins that are involved in transcriptional activation and silencing, and secondly causes relaxation of chromatin by disrupting the interactions not only between nucleosomes but also between histone tails and genomic DNA [37]. Such epigenetic regulation involves a synchronized interplay of various mechanisms that translate and integrate the signals received by modifying chromatin structure in a specific manner, which in turn regulates the gene expression. Epigenetic modifications can prime immune responses by enhancing the expression of immune-related genes and tumor-associated antigens, as well as regulating the cytokines and chemokines involved in the activation of the immune system. The development of distinct innate and adaptive immune responses can be influenced by epigenetic modifications and are therefore regulated by different epigenetic modifiers. Understanding the roles of epigenetic modifiers not only reveals the various crosslinks between epigenetic players, their inhibitors, and the immune system; but also paves way for the development of clinically significant immunotherapeutic strategies.

After undergoing positive or negative selection in the thymus, naive T cells enter the peripheral circulation, undergo clonal expansion, and differentiate into effector cells upon encounter with an antigen. The signature cytokine genes which are specific to lineage commitment such as IFN-γ in T helper 1 (Th1) cells, interleukin-4 (IL-4) in Th2 cells, and IL-17 in Th17 cells, are marked with the active histone modification H3K4 trimethylation (H3K4me3) in the appropriate lineages [38–40]. Genes encoding transcription factors like Tbx21 display a broad range of epigenetic states [41]. The Tbx21 gene encoding the Th1-specific T-bet, and the Th2-specific GATA3 gene are marked by both the activating H3K4me3 and the repressive H3K27me3 modifications in the lineages that do not express these factors [41]. H3K27me3 extensively marks IL-17 in Th1, Th2, and regulatory T (Treg) cells suggesting strong repression in these cells [41]. Chromatin modification patterns at many induced and silenced genes are stable during the short-term stimulation of resting T cells. Transcription factor E2F1, IL-2, IFN-γ, and IL-4 genes are marked with H3K4me3, H2A.Z, and other modifications even though these genes are silent in resting cells. Upon T-cell activation, the silenced genes retain positive chromatin modifications. Messi et al. observed differences in the chromatin state at IFN-γ and IL-4 promoters between naïve and memory T cells which led the authors to propose a model in which transcriptional memory, particularly epigenetic marks on chromatin, forms the basis of immunological memory [42].

Certain HDACs reverse the hyperacetylation of histones 3 and 4 at the IFN-γ promoter which controls the differentiation of naïve CD4+ T cells to Th1 and Th2 [43]. Histone hypoacetylation at the promoter region of the IL-4 gene showed reduced IL-4 transcription which impaired the Th2 differentiation of T cells [44]. Histone hyperacetylation in the proximal promoter of the IFN-γ gene was observed in Th1 cells but not Th2 cells [45]. Effector memory T cells showed polarized cytokine gene acetylation patterns in vivo, while central memory T cells had hypoacetylated cytokine genes but acquired polarized acetylation and expression after appropriate stimulation. However, hypoacetylation of the unexpressed cytokine gene did not lead to its irreversible silencing because of most Th1 and Th2 cells acetylated and expressed the alternative gene when stimulated under opposite Th polarizing conditions [42].

In Th17 cells, it has been shown that H3K4me3 and H3 acetylation mark the IL-17 and IL-17f loci [40]. Irrespective of polarizing conditions, early acetylation occurs in the

IL-4 locus which correlates with early transcription. The maintenance of acetylation depends on the cytokine/STAT (signal transducer and activator of transcription) signaling. T-bet and GATA3 also contribute to the polarized acetylated state [46]. In human and mouse nTreg cells, the forkhead box P3 (Foxp3) locus is found to be demethylated [47–49]. The unique Foxp3 promoter methylation profile in Tregs suggests that a demethylated pattern is a prerequisite for stable Foxp3 expression and suppressive phenotype [47]. Methylation of CpG residues represses Foxp3 expression, while complete demethylation is required for stable Foxp3 expression [49].

Effect of HDACi on immune cells

The various transcription factors that play an integral role in immune development include NF-κB (nuclear factor kappa-light-chain-enhancer of activated B cells), GATA-3, T-bet, and Foxp3. Interaction of HDACs with the proteins involved with these transcription factors has varying impacts on the immune system [50]. For example, downregulation of the NF-κB mediated gene transcription takes place when HDAC1 and HDAC2 bind with the NF-κB family corepressor proteins, p65 (Rel-A) and p50 (NFκB1) [51–53]. HDACs regulate the histone acetylation pattern of NF-κB, thereby maintaining the inflammatory responses mediated by cytokines IL-1β, IL-6, IL-8, and granulocyte-macrophage colony-stimulating factor (GM-CSF) [53]. Considering the important role of HDACs in the immune system, numerous HDAC inhibitors (HDACi) have been developed over the past decades to aid in various therapeutic strategies.

T cells

Histone acetylation plays an important role in CD4+ T-cell activation wherein it is involved in chromatin remodeling at the IL-2 promoter region. Histone hypoacetylation of IL-2 as well as IFN-γ promoters by HDACs and the absence of costimulation results in anergy. Inhibition of HDAC activity restores the acetylation pattern and restimulates the anergic T cells [54, 55].

In primary murine T cells, TSA (HDACi) was observed to inhibit CD4+ T-cell proliferation with a consequent increase in caspase-independent apoptosis [56]. TSA was also shown to suppress the expression of IL-2 gene and NF-κB protein levels. HDACi treatment on human T cells also showed similar inhibition of IL-2 secretion by activated T cells, with reduced expression of the IL-2 receptor marker CD25 and the T-cell activation marker CD154 (CD40L) without any effect on the expression of early activation marker CD69 [57]. TSA has been also shown to have a protective effect by blockage of peripheral blood mononuclear cellproliferation and abrogation of IFN-γ production by Th1 cells ultimately leading to apoptosis [58]. Inhibition of HDAC1 in CD4+ T cells has shown enhanced polarization to Th2 lineage. This polarization was found to be mediated by the recruitment of HDAC1 to the IL-4 gene locus in unstimulated CD4+ T cells [59]. HDACs also have a role in the polarization of CD4+ T cells toward Th17 lineage. This was observed in stimulated CD4+ T cells when HDAC inhibitors, ITF2357 and SAHA, decreased the STAT3 phosphorylation and retinoic acid receptor-related orphan receptor γT (RORγT) expression that is downstream of the IL-6R, which lead to the suppression of IL-6 and IFN-γ synthesis resulting in the reduction of Th17 cells [60].

Interestingly, HDACi VPA has also been shown to differentially modulate the surface expression and release of natural killer group 2, member D (NKG2D) ligands from various carcinoma cells and stimulate NKG2D receptor expression on γδ T cells as well as natural killer (NK) cells [61, 62]. This modulation enhances the recognition and lysis of carcinoma cells by γδ T cells and NK cells [63].

Regulatory T (Treg) cells

HDACi have shown to modulate the regulatory immunological responses as well. TSA-treated Treg cells showed increased mRNA levels of Foxp3, cytotoxic T-lymphocyte-associated protein 4 (CTLA-4), GITR, and IL-10 which augments the suppressive function of these cells [64]. Similarly, treatment of mice with TSA also increased the FOXP3 gene expression resulting in an increase in the number of Foxp3+ Treg cells and thereby augmenting their suppressive function [36]. Further studies also showed acetylation of Foxp3 by HDACs leads to its transcriptional repression thereby altering immune responses [65]. Class I HDACs, for example, deacetylate STAT3 in CD4+ T-cell leading to subsequent activation of STAT3 phosphorylation and differentiation into Foxp3+ Tregs [66]. TSA has also been found to enhance Foxp3 expression resulting in the differentiation of Treg cells from naive T cells [67].

B cells

In vitro studies have shown the impact of HDAC inhibitors on B-cell proliferation, survival, and differentiation. The effect of these inhibitors has been found to be HDAC-class dependent. The pan-HDACi vorinostat and panobinostat have been shown to modify B-cell survival, proliferation, and differentiation. On the other hand, the class-I-selective HDACi romidepsin majorly reduced B-cell survival, without much impact on its proliferation and differentiation [68].

Myeloid-derived suppressor cells

Class I HDACs play a role in the differentiation of myeloid progenitor cells to myeloid-derived suppressor cells (MDSCs), dendritic cells (DCs), and macrophages. MDSCs have been shown to express high levels of arginase 1 (ARG1), cyclooxygenase-2 (COX-2), and inducible nitric oxide synthase (iNOS) which are the key mediators of immunosuppressive activity of MDSCs. The use of entinostat, a Class I HDAC inhibitor directly targets STAT3 signaling, which impairs the T cell inhibiting activity of MDSCs by downregulating either ARG1, COX-2, or iNOS indicating that HDACs play a role in the differentiation and maintenance of MDSCs [69].

Antigen-presenting cells

HDAC inhibitors have been shown to inhibit the generation of various proinflammatory cytokines in antigen-presenting cells (APCs) such as monocytes, macrophages, and DCs. LAQ824, a pan-HDACi, modulates certain innate immune genes in lipopolysaccharide (LPS) activated APCs [70]. Additionally, LAQ824 and another pan-HDACi LBH589 also hinder the production of the antiinflammatory cytokine IL-10 in APCs. HDAC11 along with HDAC6 acts as a transcriptional activator and regulates the IL-10 secretion in APCs which play a role in the induction of immune tolerance [71, 72]. HDAC inhibitors also encourage the conversion of inflammatory macrophages into tolerogenic macrophages. They decrease Toll-Like Receptor signaling and disrupt antigen presentation [73].

Macrophages

LPS is a potent macrophage activator that triggers dramatic changes in gene expression by regulating HDAC levels in macrophages. Specific transcription regulators and chromatin regulators are responsible for activation of proinflammatory macrophages. Rapid response to an external stimulus by macrophages is associated with the permissive histone modifications on these regulators [74–77]. Particular macrophage inflammatory responses, therefore, get amplified in the presence of HDACi. The LPS-inducible expression of proatherogenic factors, COX-2 and plasminogen activator inhibitor-1 (PAI-1), was enhanced by the HDACi TSA in macrophages. Likewise, pan-HDACi, LAQ824, also amplified certain LPS responses in murine macrophages [78].

Dendritic cells

Treatment of DCs with HDACi, VPA, and entinostat (MS-275) leads to the downregulation of the costimulatory markers CD1a, CD80, CD83, and the CD54 adhesion marker impairing their differentiation and function [79]. VPA and MS-275 treatment also results in an impaired stimulatory capacity of allogeneic lymphocytes which further alters DC function [80]. These HDACi effects on DCs are mediated through alterations in the nuclear factor NFκB and interferon regulatory factor (IRF) signaling. They impair the recruitment of the RelA subunit of NF-κB and IRF-1 to the promoter of the IL-12p40 which is an HDAC-dependent gene. In an LPS/IFN-γ inflammatory model, HDACi such as TSA and SAHA inhibited the secretion of IL-12 and IL-23 cytokines of Th1 and Th17 T cells, respectively as well as selectively inhibit Th1-attracting chemokines (CXCL9, 10, 11) secreted by DCs [81]. Another HDACi, Apicidin, suppresses Th1 polarization of murine LPS-stimulated DCs derived from the bone marrow. Additionally, Apicidin attenuates the secretion of NF-κB inhibition-dependent cytokines IL-12, IL-6, tumor necrosis factor-α (TNF-α), and IFN-γ by T cells [82].

Effect of DNMTi on immune cells

The establishment and maintenance of the various patterns of DNA methylation in different cell types is mediated by DNMTs. Such methylation patterns, as well as the DNMTs associated, have been known to play a vital role in adaptive immune responses.

T cells

The DNMTi 5-Aza inhibits the proliferation of CD8+ T cells by upregulating the expression of cell cycle inhibitor p15. This leads to a shift toward CD4+ T-cell phenotype increasing the ratio of CD4+ T cells to CD8+ T cells [83]. The effect of 5-Aza on the development of CD4+ T-cell subsets Th1, Th2, and Th17 was also studied by analyzing the transcription factors T-BET1, GATA3, and RORγt, respectively [84]. 5-Aza was found to have a negative impact on Th1 differentiation, negligible effect on Th2 development, while it enhanced Th17 development [83]. Thus, treatment with 5-Aza modifies the adaptive T-cell phenotype to inhibitory T-cell phenotype.

Various deletion studies have also paved the way for better understanding of the significance of DNMTs which can be further exploited for the development of DNMTi. During T-cell development, the deletion of DNMT1 (also known as the maintenance methyltransferase) results in certain homeostatic defects as well as improper silencing of lineage-specific genes in CD4+ T-cell differentiation [85]. The deletion of DNMT3 leads to the abnormal release of cytokines and an increase in lineage plasticity [86]. Both the deletions, however, did not show any effect on the lineage specification. During CD8+ T-cell activation, the conditional deletion of DNMT1 leads to reduced effector functions, lowering the number of memory CD8+ T cells [87]. The histone methyltransferase EZH2 and DNMT1 together cause inhibition of T-cell differentiation toward a Th1 state. They also reduce the infiltration of T cells into tumors [88].

Treg cells

5-Aza treated Treg cells show decreased immune suppression with an enhanced release of IL-17. 5-Aza treated Treg cells showed the inability to suppress the proinflammatory cytokines secreted by T-effectors such as IFN-γ, IFN-α, TNF-α, RANTES, IL-6, IL-7, and IL-17 [89]. However, in a study by Stübig et al. it was reported that 5-Aza treatment increases the Treg cell number and upregulates the inhibitory cytokine (IL-10 and TGF-β) production [83].

B cells

Although the methylation by DNMTs does not affect B cell development, it restricts the activation of B cells, represses the chromatin state of plasma cells, and regulates the differentiation of plasma cells [90]. Along with DNMT, HMTs also have an important role in B cell immune responses. Inhibition of the EZH2 decreases the number of memory B cells and results in reduced antibody responses [91].

Macrophages

Macrophages have two distinct polarized states which include classically activated (M1) macrophages and alternatively activated (M2) macrophages that are driven by cytokine repertoire of Th1 and Th2 cells, respectively. In the proinflammatory tumor microenvironment, immune cells like monocytes differentiate into tumor-suppressive macrophages (M1) and tumor-promoting macrophages (M2) also called tumor-associated macrophages [92]. DNMTs decrease TNF production and increase the methylation of PPARγ1 promoter region which contributes to the polarization of macrophages [93]. 5-Aza as well as decitabine decrease the M1 expression and increase the M2 expression [94]. 5-Aza decreases iNOS as well as TNF expression and NO production associated with M1 polarization but increases the expression of M2 markers ARG1 and FIZZ1 [95–97]. Similarly, decitabine reduced the expression of M1 markers such as iNOS, TNF, IL1α, IL1β, and IL6 as well as various chemokines (CCL2, CCL5, and CCL9); whereas the expression of M2 marker ARG1 was increased [98, 99]. It was also found that decitabine in combination with TSA (HDACi) had an enhanced effect on the expression of NOS2, ARG1, and CD206 resulting in M2 polarization [99].

Dendritic cells

Mature DCs in the presence of 5-Aza upregulate the expression of the surface markers CD40 and CD86 which are required for the priming and activation of T cells, respectively. The treatment also lowers the levels of cytokines secreted by DCs, predominantly IL-10 and IL-27. 5-Aza treatment may also be altering T cell polarization by favoring Th17 development [100].

Hence, the various epigenetic modifiers influence the development and the effector functions of different immune cells (Fig. 1).

Clinical significance of epigenetic modifiers

Global analyses of the specific mutations observed in certain disorders such as lymphoid malignancies have shown that such mutations influence the function of epigenetic modifying enzymes in addition to the function of transcription factors, cytokine signaling and antigen receptors involved [101]. This has led to the development of various epigenetic modifiers which include HDACi and DNMTi for the pharmacological treatment of different disorders. Different epigenetic modifiers are currently being used in clinical trials either as monotherapy or in combination with other treatment modalities (Tables 1 and 2).

HDACi therapy

The role of HDACi for therapeutic purposes was first studied in vitro for tumor eradication on different cell lines. The treatment of HDACi TSA, sodium butyrate, and trapoxin A on mouse plasmacytoma cells, human neuroblastoma tumor cells, and acute myeloid leukemic (AML) cells showed the upregulation of MHC class I and II proteins, costimulatory antigens

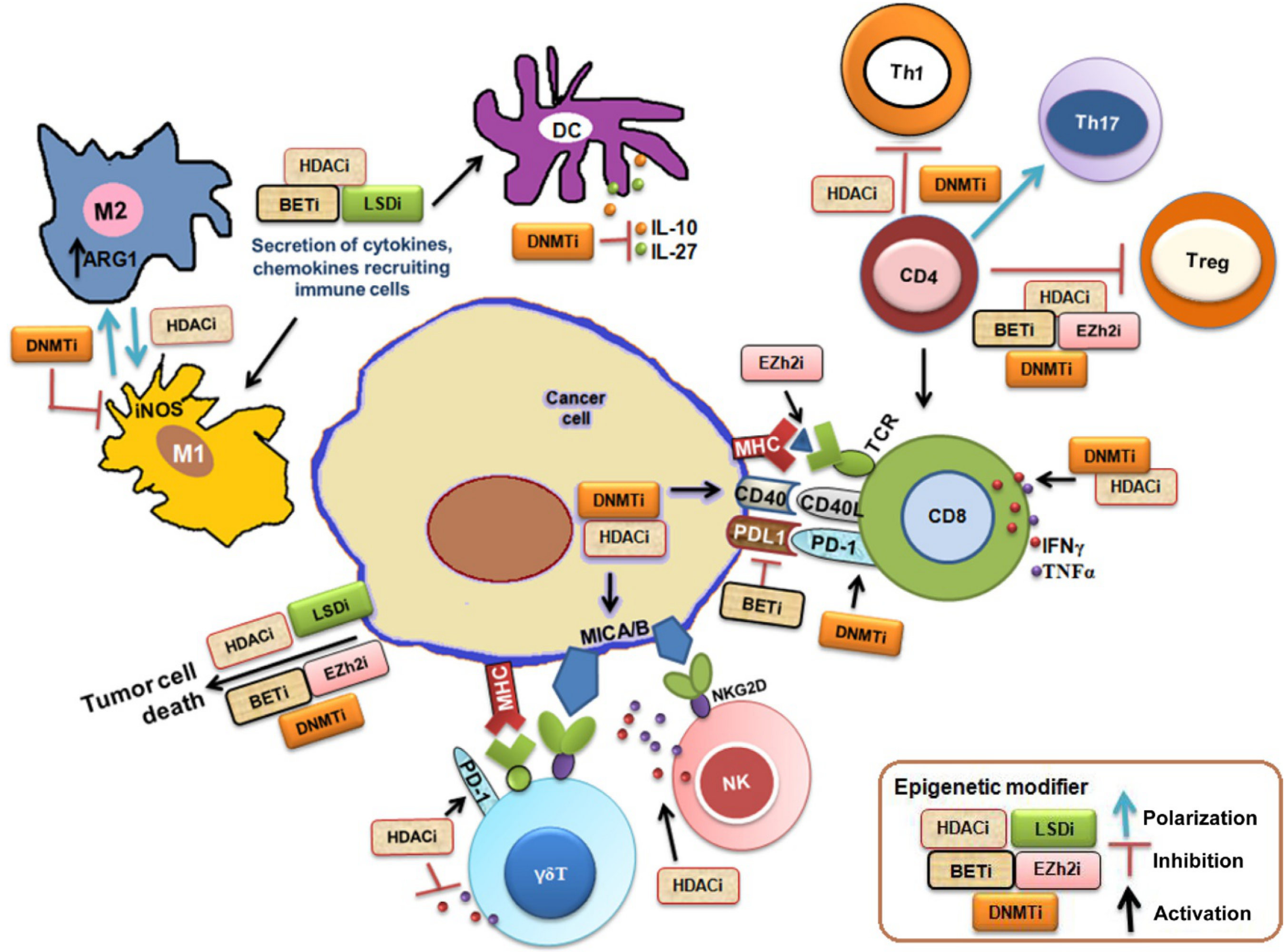

M2
↑ARG1
HDACi
BETi
LSDi
Secretion of cytokines, chemokines recruiting immune cells
DC
DNMTi
IL-10
IL-27
Th1
Th17
HDACi
DNMTi
CD4
Treg
HDACi
BETi
EZh2i
DNMTi
DNMTi
HDACi
iNOS
M1
Cancer cell
EZh2i
MHC
TCR
DNMTi
HDACi
CD40
CD40L
PDL1
PD-1
CD8
DNMTi
HDACi
IFNγ
TNFα
BETi
DNMTi
MICA/B
LSDi
HDACi
EZh2i
BETi
DNMTi
Tumor cell death
MHC
NKG2D
PD-1
NK
HDACi
γδT
HDACi
Epigenetic modifier
HDACi
LSDi
BETi
EZh2i
DNMTi
Polarization
Inhibition
Activation

Fig. 1 Differential effect of epigenetic modifiers on immune cells. Epigenetic modifiers like HDACi, LSDi, BETi, EZH2i, and DNMTi bring about tumor cell death through induction of apoptosis and also regulate the balance of immunosuppression/immune activation which is important for tumor immunotherapy. Epigenetic modifiers (1) inhibit the regulatory T cells, Th1, and M1 macrophages; (2) promote Th17 and M2 polarization; and (3) activate NK cells, CD8, and γδ T cells through upregulation of costimulatory molecules on tumor cells and secretion of cytokines and chemokines by DCs and M1 macrophages.

Table 1 List of clinical trials summarizing the use of epigenetic modifiers as monotherapy against various diseases.

Sr. No.	Clinical trial identifier	Phase	Condition	Epigenetic drug	Drug target
1	NCT00274651	II	Cutaneous T-cell lymphoma, peripheral T-cell lymphoma, non-Hodgkin's lymphoma	Belinostat	HDAC
2	NCT00412997	I	Cutaneous T-cell lymphoma	Panobinostat	HDAC
3	NCT00699296	II	Cutaneous T-cell lymphoma, adult T-cell leukemia/lymphoma	Panobinostat	HDAC
4	NCT00490776	II/III	Cutaneous T-cell lymphoma	Panobinostat	HDAC
5	NCT00425555	II/III	Cutaneous T-cell lymphoma	Panobinostat	HDAC
6	NCT01680094	I/II	HIV infection	Panobinostat	HDAC
7	NCT01486277	II	Cutaneous T-cell lymphoma	Quisinostat	HDAC
8	NCT00007345	II	Cutaneous T-cell lymphoma, peripheral T-cell lymphoma	Romidepsin	HDAC
9	NCT01433731	I	Cutaneous T-cell lymphoma	SHAPE	HDAC
10	NCT02213861	II	Cutaneous T-cell lymphoma	SHAPE	HDAC
11	NCT02944812	II	Peripheral T-cell lymphoma	Chidamide	HDAC
12	NCT02124083	I/II	Neimann-Pick disease	Vorinostat	HDAC
13	NCT01933594	I/II	HIV infection	Romidepsin	HDAC
14	NCT02953301	II	Mycosis fungoides, sezary syndrome, cutaneous T-cell lymphoma	Resminostat	HDAC
15	NCT02759601	I/II	Hepatocellular carcinoma	Tefinostat	HDAC
16	NCT00212316	II	Huntington's disease	Sodium phenylbutyrate	HDAC
17	NCT00107770	I/II	Amyotrophic lateral sclerosis	Sodium phenylbutyrate	HDAC
18	NCT01589809	II	Neurodegenerative disorders	Nicotinamide	HDAC
19	NCT02034123	I	Relapsed/refractory small cell lung carcinoma	GSK2879552	HDM
20	NCT01430455	IV	Bipolar disorder I or II	Tranylcypromine	HDM
21	INCB059872	I	Relapsed Ewing sarcoma	INCB059872	HDM
22	NCT03456726	II	Relapsed or refractory B-cell non-Hodgkin's lymphoma	Tazemetostat	HMT
23	NCT02900651	I/II	Diffuse large B-cell lymphoma	MAK683	HMT
24	NCT02082977	I	Cancer	GSK2816126	HMT
25	NCT02395601	I	B-cell lymphoma	CPI-1205	HMT
26	NCT01897571	I/II	B-cell lymphomas (phase 1), advanced solid tumors (phase1), diffuse large B-cell lymphoma (phase 2), follicular lymphoma (phase 2)	Tazemetostat	EZH2

Continued

Table 1 List of clinical trials summarizing the use of epigenetic modifiers as monotherapy against various diseases—cont'd

Sr. No.	Clinical trial identifier	Phase	Condition	Epigenetic drug	Drug target
27	NCT02601937	I	Rhabdoid tumors, INI1-negative tumors, synovial sarcoma, malignant rhabdoid tumor of ovary, stage II breast cancer, stage III breast cancer, invasive breast cancer	Tazemetostat	EZH2
28	NCT02601950	II	Malignant rhabdoid tumors, Rhabdoid tumors of the kidney, atypical teratoid rhabdoid tumors, selected tumors with rhabdoid features, synovial sarcoma, INI1-negative tumors, malignant rhabdoid tumor of ovary, renal medullary carcinoma, epithelioid sarcoma, any solid tumor with an EZH2 GOF mutation	Tazemetostat	EZH2
29	NCT02920008	III	Acute myeloid leukemia	Guadecitabine	DNMT
30	NCT02423057	I	Neoplasms, solid tumors	TdCyd	DNMT

HDAC, histone deacetylase; *EZH2*, enhancer of zeste homolog 2; *HMD*, histone demethylase; *HMT*, histone methyl transferase; *DNMT*, DNA methyltransferase.

(CD40, CD80, and CD86), and adhesion molecules (intercellular adhesion molecule-1 (ICAM-1)) [68, 102]. Thus, HDACi were shown to play a role in tumor eradication [102]. Similarly HDACi TSA, SAHA, VPA, and sodium butyrate increase the MHC class I related chain A (MICA) and chain B (MICB) expression on tumor cells which are ligands for NKG2D on NK cells, γδ T cells, and CD8+ T cells [103, 104]. This activation ultimately results in the immune-targeted killing of tumor cells. HDACi TSA, as well as RNAi-mediated inhibition of HDAC11, has been found to enhance tumor cell cytotoxicity [105, 106]. HDACi such as TSA and depsipeptide have been used in cancer vaccination and have been found to increase cancer/germ line family of antigens, cancer vaccine target expression, as well as MHC and other costimulatory molecules on tumors [107]. TSA has also been found to augment the expression of CD40, MHC class II, and B7-1/2 antigens on B16 melanoma cells which activates NK cells and IFN-γ-secreting CD8+ T cells [108]. Surprisingly, TSA has been shown to assist tumor escape from the immune system by reducing the presentation and destruction of tumor cells by CD8+ T cells [109].

HDACs have also been targeted in various inflammatory conditions based on their crucial role during T-cell differentiation and effector functions. The use of HDACi under such conditions has shown suppression of immune responses [110]. In inflammatory

Table 2 List of clinical trials summarizing the use of epigenetic modifiers as combinatorial therapeutic agents against various diseases.

Sr. No.	Clinical trial identifier	Phase	Condition	Epigenetic drug	Combination drug (chemotherapy drug/ immune checkpoint inhibitor)	Drug target
1	NCT00828854	II	Estrogen receptor positive breast cancer	Entinostat		HDAC
2	NCT02513901	I/II	Chronic HIV infections	Chidamide		HDAC
3	NCT02718066	I/II	Melanoma, renal cell carcinoma, non-small-cell lung cancer	Chidamide (HBI-8000)	Nivolumab	HDAC PD-1
4	NCT01023308	III	Multiple myeloma	Panobinostat	Bortezomib	HDAC Proteasome
5	NCT02032810	I	Melanoma, skin cancer	Panobinostat	Ipilimumab	HDAC CTLA-4
6	NCT02890069	I	Colorectal cancer, non-small-cell lung adenocarcinoma, triple negative breast cancer, renal cell carcinoma	Panobinostat	PDR001, LCL161, Everolimus, QBM076, HDM201	HDAC PD-1 Inhibitor of apoptosis (IAP) proteins mTOR kinase CXCR2 HDM2 protein
7	NCT01282476	II	Diffuse large B cell lymphoma	Panobinostat	Rituximab	HDAC CD20
8	NCT00567879	I/II	Breast cancer	Panobinostat	Trastuzumab	HDAC Human epidermal growth factor receptor 3

Continued

Table 2 List of clinical trials summarizing the use of epigenetic modifiers as combinatorial therapeutic agents against various diseases—cont'd

Sr. No.	Clinical trial identifier	Phase	Condition	Epigenetic drug	Combination drug (chemotherapy drug/ immune checkpoint inhibitor)	Drug target
9	NCT03250273	II	Previously treated unresectable or metastatic cholangiocarcinoma and pancreatic cancer, cholangiocarcinoma, pancreatic cancer, metastatic pancreatic cancer	Entinostat	Nivolumab	HDAC PD-1
10	NCT02453620	I	Breast adenocarcinoma, HER2/Neu negative invasive breast carcinoma, metastatic breast carcinoma, metastatic malignant solid neoplasm, stages III, IIIA, IIIB, IIIC breast cancer AJCC v7, stage IV breast cancer AJCC v6 and v7, unresectable solid neoplasm	Entinostat	Ipilimumab, Nivolumab	HDAC CTLA-4 PD-1
11	NCT03179930	II	Relapsed and refractory lymphomas	Entinostat	Pembrolizumab	HDAC PD-1
12	NCT02936752	I	Blasts 21%–30% of bone marrow nucleated cells, myelodysplastic syndrome, previously treated myelodysplastic syndrome	Entinostat	Pembrolizumab	HDAC PD-1

13	NCT02909452	I	Neoplasms, glandular neoplasms, epithelial neoplasms, neoplasms by histologic type, bronchial neoplasms, lung neoplasms, respiratory tract neoplasms, thoracic neoplasms, digestive system neoplasms, endocrine gland neoplasms, non-small-cell lung carcinoma, lung diseases, breast diseases, renal neoplasm, solid tumors	Entinostat	Pembrolizumab	HDAC PD-1
14	NCT03024437	I/II	Metastatic cancer, renal cancer	Entinostat	Atezolizumab, Bevacizumab	HDAC PD-L1 VEGF-A
15	NCT02708680	I/II	Breast cancer	Entinostat	Atezolizumab	HDAC PD-L1
16	NCT02915523	I/II	Epithelial ovarian cancer, peritoneal cancer, fallopian tube cancer	Entinostat	Avelumab	HDAC PD-L2
17	NCT02697630	II	Metastatic uveal melanoma	Entinostat	Pembrolizumab	HDAC PD-1
18	NCT02437136	I/II	Non-small-cell lung cancer, melanoma, mismatch repair-proficient colorectal cancer	Entinostat	Pembrolizumab	HDAC PD-1
19	NCT02820961	I	Breast cancer, estrogen receptor positive breast cancer	Entinostat	Exemestane	HDAC Estrogen receptors
20	NCT02805660	I/II	Advanced cancer	Mocetinostat	Durvalumab	HDAC PD-L1

Continued

Table 2 List of clinical trials summarizing the use of epigenetic modifiers as combinatorial therapeutic agents against various diseases—cont'd

Sr. No.	Clinical trial identifier	Phase	Condition	Epigenetic drug	Combination drug (chemotherapy drug/ immune checkpoint inhibitor)	Drug target
21	NCT02954991	II	Non-small-cell lung carcinoma	Mocetinostat	Glesatinib, Sitravatinib, Nivolumab	HDAC Receptor tyrosine kinases PD-1
22	NCT02993991	I	Head and neck squamous cell carcinoma, mouth squamous cell carcinoma, resectable squamous cell carcinoma of oral cavity	Mocetinostat	Durvalumab	HDAC PD-L1
23	NCT01238692	II	Diffuse large B cell lymphoma	LBH589	Rituximab	HDAC CD20
24	NCT02512172	I	Colorectal cancer	Romidepsin + Oral CC-486 (oral azacitidine)	Pembrolizumab (MK-3475)	HDAC DNMT PD-1
25	NCT02538510	I/II	Head and neck squamous cell carcinoma, recurrent nasal cavity and paranasal sinus squamous cell carcinoma, recurrent nasopharynx carcinoma, recurrent salivary gland carcinoma, squamous cell carcinoma metastatic in the neck with occult primary, stage iii major salivary gland carcinoma, stage III nasal cavity	Vorinostat	Pembrolizumab	HDAC PD-1

			and paranasal sinus squamous cell carcinoma, stages III, IV nasopharyngeal carcinoma, stages IVA, IVB, IVC major salivary gland carcinoma, stages IVA, IVB, IVC nasal cavity and paranasal sinus squamous cell carcinoma, squamous cell carcinoma			
26	NCT02707900	I	HIV-1 infection	Vorinostat	AGS-004	HDAC Dendritic cells
27	NCT02638090	I/II	Lung cancer, non-small-cell lung cancer	Vorinostat	Pembrolizumab	HDAC PD-1
28	NCT02619253	I	Renal cell carcinoma, urinary bladder neoplasms	Vorinostat	Pembrolizumab	HDAC PD-1
29	NCT02395627	II	Breast neoplasms	Vorinostat	Pembrolizumab, Tamoxifen	HDAC PD-1 Estrogen receptors
30	NCT00667615	I/II	Hodgkin's disease, lymphoma	Vorinostat	Cyclophosphamide, Etoposide, Prednisone and Rituximab	HDAC CD20 DNA topoisomerase II Others
31	NCT00720876	II	Lymphoma	Vorinostat	Rituximab	HDAC CD20
32	NCT00972478	I/II	Ann Arbor stage II, III, IV non-Hodgkin lymphoma	Vorinostat	Rituximab, Cyclophosphamide, Doxorubicin hydrochloride, Vincristine sulfate, and Prednisone	HDAC CD20 DNA topoisomerase II Others

Continued

Table 2 List of clinical trials summarizing the use of epigenetic modifiers as combinatorial therapeutic agents against various diseases—cont'd

Sr. No.	Clinical trial identifier	Phase	Condition	Epigenetic drug	Combination drug (chemotherapy drug/ immune checkpoint inhibitor)	Drug target
33	NCT00764517	II	Recurrent and refractory B-cell non-Hodgkin lymphoma, recurrent chronic lymphocytic leukemia, recurrent indolent adult non-Hodgkin lymphoma	Vorinostat	Cladribine and Rituximab	HDAC CD20 Other
34	NCT00918723	I/II	Chronic lymphocytic leukemia, stages I, II, III chronic lymphocytic leukemia, stages I, II, III small lymphocytic lymphoma	Vorinostat	Fludarabine phosphate, Cyclophosphamide, and rituximab	HDAC CD20 DNA synthesis
35	NCT01194427	II	Stage I breast cancer, stage II breast cancer, stage III breast cancer, invasive breast cancer	Vorinostat	Tamoxifen	HDAC Estrogen receptors
36	NCT00365599	II	Breast cancer	Vorinostat	Tamoxifen	HDAC Estrogen receptors
37	NCT00616967	II	Breast cancer	Vorinostat	Carboplatin and Paclitaxel albumin-stabilized nanoparticle formulation	HDAC DNA replication Microtubules
38	NCT00368875	I/II	Male breast cancer, stage IIIB breast cancer, stage IIIC breast cancer, stage IV breast cancer	Vorinostat	Paclitaxel	HDAC Microtubules
39	NCT00258349	I/II	Breast cancer, male breast cancer, recurrent breast cancer, stage IIIB breast cancer, stage IIIC breast cancer, stage IV breast cancer	Vorinostat	Trastuzumab	HDAC Human epidermal growth factor receptor 2

40	NCT01010854	II	Breast cancer	Valproic acid	FEC100	HDAC Nucleic acid synthesis
41	NCT00227266	II	Spinal muscular atrophy	Valproic acid	Levocarnitine	HDAC Mitochondrial membrane carrier
42	NCT02635061	I	Non-small-cell lung cancer	ACY-241	Nivolumab	HDAC PD-1
43	NCT01686165	II	Anaplastic large cell lymphoma, recurrent adult diffuse large cell lymphoma, recurrent mantle cell lymphoma	Belinostat	Rituximab, Yttrium Y 90 ibritumomab tiuxetan	HDAC CD20
44	NCT02948075	II	Ovarian cancer	Quisinostat (JNJ-26481585)	Paclitaxel and Carboplatin	HDAC DNA replication Microtubules
45	NCT03161223	I/II	T cell lymphoma	Romidepsin +Azacitidine	Durvalumab and Pralatrexate	HDAC DNMT PD-L1 Dihydrofolate reductase
46	NCT03220477	I	Lung cancer	Mocetinostat +Guadecitabine	Pembrolizumab	HDAC DNMT PD-1
47	NCT03151408	III	Acute myeloid leukemia	Pracinostat +Azacitidine		HDAC DNMT
48	NCT02220842	I	Lymphoma	Tazemetostat	Atezolizumab and Obinutuzumab	EZH2 PD-L1 CD20

Continued

Table 2 List of clinical trials summarizing the use of epigenetic modifiers as combinatorial therapeutic agents against various diseases—cont'd

Sr. No.	Clinical trial identifier	Phase	Condition	Epigenetic drug	Combination drug (chemotherapy drug/ immune checkpoint inhibitor)	Drug target
49	NCT02397720	II	Acute bilineal leukemia, acute biphenotypic leukemia, acute myeloid leukemia arising from previous myelodysplastic syndrome, chronic myelomonocytic leukemia, myelodysplastic syndrome; recurrent, refractory, secondary and therapy-related acute myeloid leukemia	Azacitidine	Ipilimumab and Nivolumab	DNMT CTLA-4 PD-1
50	NCT02260440	II	Metastatic colorectal cancer	Azacitidine	Pembrolizumab	DNMT PD-1
51	NCT02546986	II	Non-small-cell lung carcinoma	CC-486 (oral azacitidine)	Pembrolizumab	DNMT PD-1
52	NCT02508870	I	Myelodysplastic syndromes	Azacitidine	Atezolizumab	DNMT PD-L1
53	NCT02951156	III	Diffuse large B-cell lymphoma	Azacitidine	Avelumab, Utomilumab, Rituximab, Bendamustine, Gemcitabine and Oxaliplatin	DNMT PD-L1 CD137 (4-1BB) CD20 DNA synthesis
54	NCT02530463	II	Leukemia, myelodysplastic syndrome	Azacitidine	Nivolumab and Ipilimumab	DNMT PD-1 CTLA-4

55	NCT02399917	II	Acute biphenotypic leukemia, acute myeloid leukemia arising from previous myelodysplastic syndrome, recurrent and refractory acute myeloid leukemia	Azacitidine	Lirilumab	DNMT KIR2DL1/2L3
56	NCT02599649	II	Leukemia	Azacitidine	Lirilumab and Nivolumab	DNMT KIR2DL1/2L3 PD-1
57	NCT02816021	II	Melanoma and other malignant neoplasms of skin, metastatic melanoma	Azacitidine	Pembrolizumab	DNMT PD-1
58	NCT01928576	II	Non-small lung cancer	Azacitidine +Entinostat	Nivolumab	DNMT HDAC PD-1
59	NCT02811497	II	Microsatellite stable colorectal carcinoma, platinum resistant epithelial ovarian cancer type II, estrogen receptor positive and HER2 negative breast cancer	Azacitidine	Durvalumab	DNMT PD-L1
60	NCT02281084	II	Myelodysplastic syndromes	CC-486 (oral azacitidine)	Durvalumab	DNMT PD-L1
61	NCT02775903	II	Acute myeloid leukemia, myelodysplastic syndromes	Azacitidine	Durvalumab	DNMT PD-L1
62	NCT02953561	I/II	Acute myeloid leukemia arising from previous myelodysplastic syndrome, recurrent acute myeloid leukemia, refractory acute myeloid leukemia	Azacitidine	Avelumab	DNMT PD-L1

Continued

Table 2 List of clinical trials summarizing the use of epigenetic modifiers as combinatorial therapeutic agents against various diseases—cont'd

Sr. No.	Clinical trial identifier	Phase	Condition	Epigenetic drug	Combination drug (chemotherapy drug/ immune checkpoint inhibitor)	Drug target
63	NCT03019003	I/II	Head and neck cancer	Azacitidine	Durvalumab and Tremelimumab	DNMT PD-L1 CTLA-4
64	NCT02117219	I	Myelodysplastic syndrome	Azacitidine	Durvalumab and Tremelimumab	DNMT PD-L1 CTLA-5
65	NCT03092674	II/III	Acute myeloid leukemia, myelodysplastic syndrome, myelodysplastic syndrome with excess blasts-2, untreated adult acute myeloid leukemia	Azacitidine +Decitabine	Cytarabine, Midostaurin and Nivolumab	DNMT DNA synthesis FMS-like tyrosine kinase 3 PD-1
66	NCT03066648	I	Leukemia, myeloid leukemia, acute myeloid leukemia, myelodysplastic syndromes, preleukemia, bone marrow diseases, hematologic diseases	Decitabine	PDR001 and MBG453	DNMT PD-1 TIM-3
67	NCT02890329	I	Blasts 5% or more of bone marrow nucleated cells, hematopoietic cell transplant recipient, myelodysplastic syndrome, previously treated myelodysplastic syndrome, recurrent acute myeloid leukemia, recurrent acute myeloid leukemia with	Decitabine	Ipilimumab	DNMT CTLA-4

			myelodysplasia-related changes, refractory acute myeloid leukemia, secondary acute myeloid leukemia, untreated adult acute myeloid leukemia			
68	NCT02795923	II	Lung cancer, non-small-cell lung cancer	Decitabine	Nivolumab and Tetrahydrouridine	DNMT PD-1 Cytidine deaminase
69	NCT02664181	II	Lung cancer, non-small-cell lung cancer	Decitabine	Nivolumab and Tetrahydrouridine	DNMT PD-1 Cytidine deaminase
70	NCT03233724	I/II	Lung cancer, non-small-cell lung cancer	Decitabine	Pembrolizumab and Tetrahydrouridine	DNMT PD-1 Cytidine deaminase
71	NCT02608437	I	Metastatic melanoma	Guadecitabine	Ipilimumab	DNMT CTLA-4
72	NCT02935361	I/II	Chronic myelomonocytic leukemia, myelodysplastic syndrome, recurrent acute myeloid leukemia with myelodysplasia-related changes	Guadecitabine	Atezolizumab	DNMT PD-L1
73	NCT03206047	I/II	Platinum-resistant fallopian tube carcinoma, platinum-resistant ovarian carcinoma, platinum-resistant primary peritoneal carcinoma, recurrent fallopian tube carcinoma, recurrent ovarian carcinoma, recurrent primary peritoneal carcinoma	Guadecitabine	Atezolizumab CDX-1401	DNMT PD-L1 Endocytic DC receptor

Continued

Table 2 List of clinical trials summarizing the use of epigenetic modifiers as combinatorial therapeutic agents against various diseases—cont'd

Sr. No.	Clinical trial identifier	Phase	Condition	Epigenetic drug	Combination drug (chemotherapy drug/ immune checkpoint inhibitor)	Drug target
74	NCT02892318	I	Acute myeloid leukemia	Guadecitabine	Atezolizumab	DNMT PD-L1
75	NCT03179943	II	Urothelial carcinoma	Guadecitabine	Atezolizumab	DNMT PD-L1
76	NCT03257761	I	Hepatocellular, pancreatic, cholangiocarcinoma, gallbladder	Guadecitabine	Durvalumab	DNMT PD-L2
77	NCT03292172	I	Advanced ovarian cancer, triple negative breast cancer	RO6870810	Atezolizumab	BET PD-L1

HDAC, histone deacetylase; *EZH2*, enhancer of zeste homolog 2; *DNMT*, DNA methyltransferase; *PD1*, programmed death protein 1; *PDL-1/2*, programmed death ligand-1/2; *BET*, bromodomain and extra-terminal motif proteins; *CTLA-4*, cytotoxic T lymphocyte-associated antigen-4; *TIM3*, T-cell immunoglobulin and mucin-domain containing-3; *CXCR2*, C-X-C motif chemokine receptor 2; *HDM2*, human homologue of mouse double minute 2 homolog.

conditions, HDAC inhibitors have been shown to induce regulatory T-cells generation or stabilization of regulatory T cells in inflammatory conditions thereby maintaining the tolerance. Due to this property of inhibition of inflammatory conditions, HDAC inhibitors have shown promising results in experimental colitis [111, 112].

DNMTi therapy

DNMTi are being widely exploited for cancer therapy since they boost the body's immune responses by enhancing the presentation and expression of cancer antigens as well as by increasing the IFN response [113]. The DNMTi 5-Aza and decitabine were among the first epigenetic modifiers to be tested and approved by the US FDA for therapeutic purposes [114]. They have been approved for the treatment of myelodysplastic syndrome (MDS). However, they have been associated with cytotoxicity problems wherein they decrease the growth and proliferation of myeloid-derived blood cells [115].

In phase II clinical trial with DNMTi MG98 for DNMT1 inhibition in patients with metastatic renal carcinoma, there was no clear antitumor activity observed [116]. A phase I trial focuses on the use of decitabine for targeting aberrant DNA hypermethylation in chronic lymphocytic leukemia (CLL) and non-Hodgkin's lymphoma (NHL) [117]. Currently, an alternative strategy has been developed in which treatment with 5-Aza is used to prime cancerous cells such as diffuse large B-cell lymphoma (DLBCL) before the administration of chemotherapy. Here, DNMTi treatment may be playing a role in the demethylation and reexpression of genes thereby modulating the sensitivity of lymphoma cells to chemotherapy. A phase I trial of the same combination was found to be tolerable with manageable toxicity, even in older patients [118]. Similarly, decitabine is being tested in a phase II trial for the treatment of patients with follicular thyroid cancer or metastatic papillary thyroid cancer unresponsive to iodine [119].

Other epigenetic modifiers in therapy

Histone methyltransferase and demethylase enzymes are also being exploited clinically for various malignancies. H3K27 trimethylation repression is one of the epigenetic mechanisms that have been observed in different tumors. The repression mediated by H3K27 trimethylation occurs by the activity of the histone methyltransferase, EZH2. EZH2 has been associated with increased proliferation, neoplastic transformation, and is overexpressed in several cancer types [120–124]. Inhibition of EZH2 influences the immunogenicity of cancer cells directly by modulation of the tumor microenvironment or by affecting the antigen presentation process. EZH2 inhibitors, along with DNMT inhibitors have been used to stimulate "cold" state tumor microenvironment to "hot" state. These inhibitors together enhance the expression of the Th1-type chemokines CXCL9 and CXCL10, resulting in an increase in the release of T cells in the tumor microenvironment [113].

The first selective inhibitor of HMT DOT1L, EPZ004777 was found to selectively destroy MLL-rearranged leukemia cells in vitro with negligible effect on non-MLL translocated cells and better survival in mice with MLL-rearranged leukemia [125]. In spite of these positive results, EPZ004777 was found to have poor pharmacokinetic properties which made it unsuitable for clinical development [126]. Thus, the second generation of DOT1L inhibitors was developed with better pharmacokinetic properties which included one of the novel derivatives of EPZ004777, EPZ-5676 (Pinometostat), with enhanced activity against DOT1L [127]. This inhibitor reduced the proliferation of leukemic cell lines with MLL translocation, with negligible effect on cells without the translocation [125, 127]. Although EPZ-5676 showed low oral bioavailability, continuous infusions were surprisingly well tolerated in rat xenograft model of MLL-rearranged leukemia [126, 128]. It has also been found to have synergistic effect with common AML standard care drugs which include daunorubicin, cytarabine, and DNMTi azacitidine [129].

Combination therapy involving epigenetic modifiers

Advanced research on the fine mechanisms regulating the crosstalk of immune cells with multi-faceted tumors has led to the development of several novel immunotherapeutic strategies for the treatment of cancer. Modulation of immune responses by regulating the expression of different antigens through epigenetic mechanisms is well studied. Epigenetic modifiers have been shown to increase the expression of antigens such as MHC I and MHC II molecules on the tumor cells which leads to increased immune response between tumor and immune cells [130]. Targeting these immunological synapses has shown notable clinical efficacy in different tumor types [131, 132]. Despite the antitumor potential of the epigenetic modifiers, the clinical responses and, most importantly, long-term survival are achieved in only a small fraction of patients when used as monotherapy. Further, epigenetic modifiers have also been shown to have a negative impact on the effector functions of different immune systems and result in resistance to treatment [133]. To overcome such negative impacts of epigenetic drugs, combination therapies involving epigenetic drugs along with other modalities such as immunotherapy, chemotherapy, and radiation are being explored.

Epigenetic modifiers and immunotherapy

Various in vitro and in vivo studies involving the combination of epigenetic modifiers and immune checkpoint inhibitors for therapeutic purposes has shown promising results [134, 135]. Monoclonal antibodies against different molecules have revolutionized the field of immunotherapeutics. Modulating immune inhibitory pathways by monoclonal antibodies against immune checkpoints, such as programmed cell death protein-1 (PD-1) or CTLA-4 receptors have been considered as an important breakthrough in the field of immune therapeutic approaches [136, 137]. Although these immune checkpoint blockades have improved patient outcomes in both solid and hematological malignancies, the main concern of these therapies is the immune escape of the tumors [133]. Recent studies by our group and others have shown that a combinatorial approach using immune checkpoint blockade and epigenetic drugs has enhanced antitumor immune responses mediated by γδ T cells and other immune cells [22, 138]. Combination of DNMTi azacitidine and anti-PD-1 monoclonal antibody Nivolumab is in phase II trial for lung cancer. Combinatorial administration of the HDACi vorinostat and the anti-PD-1 blockade antibody pembrolizumab are in phase I/II trial for non-small-cell lung cancer [134]. In a recent study by Kim et al., it was seen that the treatment of anti-PD-1 and anti-CTLA-4 antibodies in tumor-bearing mice was unable to eradicate the tumors [139]. However, treatment of anti-PD-1 and anti-CTLA-4 in combination with HDACi entinostat improved treatment outcomes, curing more than 80% of the tumor-bearing mice [139]. Better understanding of the molecular mechanisms of how epigenetic drugs modulate the immune response would contribute to their clinical development as anticancer agents more efficiently. Besides being used in combinatorial approaches along with immune checkpoint blockades, HDACi panobinostat in combination with adoptive T cell transfer in the B16 melanoma model has also shown promising results [140]. DNMTi decitabine, when used in combination with anti-CTLA-4 therapy, increases the cytokine production in NK cells and CD8+ cytotoxic T cells [141]. Decitabine has also shown a dose-dependent upregulation of PD-L1, PD-L2, and PD-1 expression in certain leukemias. Combined PD-1 blockade and DNMT inhibitor therapy have been used to illustrate that epigenetic therapy may be used as a sensitizer to immune checkpoint inhibitors [142].

Epigenetic modifiers and chemotherapeutics

There have been various studies to analyze the combined effect of epigenetic modifiers and chemotherapeutic drugs. One of the age-old problems faced with conventional chemotherapy is the development of multidrug resistance. The administration of HDACi and/or DNMTi before chemotherapy could be used as a potential strategy to overcome such resistance. Numerous preclinical studies, as well as clinical trials, have been carried out to correlate the effects of such combination therapy with better patient survival. The anticancer activity of paclitaxel was found to increase in vivo when combined with HDACi TSA and in vitro with VPA [143, 144]. In patients with advanced solid tumors enrolled for phase I clinical trials, HDACi vorinostat in combination with the chemotherapeutic drugs paclitaxel and carboplatin showed promising anticancer potential [145]. Similar results were seen when HDACi tacedinaline (CI 994) was used in combination with paclitaxel and carboplatin for the treatment of solid tumors [146]. Metastatic non-small-cell lung cancer patients showed enhanced clinical outcomes when treated with HDACi SAHA combined with paclitaxel and carboplatin [147]. Combination of HDACi entinostat with an aromatase inhibitor also showed promising outcomes in breast cancer patients [148]. These clinical trials highlighted the feasibility of such combination therapy in various cancers with promising outcomes.

HDAC inhibitors have also been shown to improve the cytotoxicity of other traditional chemotherapeutic drugs like 5-fluorouracil (5-FU) in human cancerous cells. A phase I clinical trial with metastatic colorectal cancer patients showed that treatment of HDACi sodium phenylbutyrate along with 5-FU was well tolerated and could potentially stabilize the disease [149, 150]. Similarly, in phase I clinical trial enrolling patients with solid tumors, the combination of HDACi tacedinaline with chemotherapeutic drug capecitabine showed addictive antitumor effect [151]. These clinical trials put forth the ability of epigenetic modifiers in stimulating the anticancer potential of chemotherapeutic drugs.

Epigenetic modifiers and radiotherapy

Ionizing radiation which is widely used in the treatment of numerous carcinomas causes cell death by inducing DNA double-strand breaks (DSBs). HDACi have been found to promote the sensitivity of cancer cells to radiation [152, 153]. Such activity of HDACi has been seen in vitro in different cancer cell lines including melanoma, colon, non-small cell lung (NSCLC), prostate, and others [154]. HDACi sodium butyrate was the first to be used in combination with radiation to sensitize human colon cancer cells [155]. There have been preclinical studies in melanoma, NSCLC, head and neck squamous, glioblastoma multiforme, colorectal, and breast cancer [156]. HDACi in combination with radiotherapy was reported for the treatment of anaplastic thyroid carcinoma. Along with radiation, the HDACi VPA was used in combination with cisplatin, doxorubicin, and surgery for successful treatment of the patient [157]. Currently, numerous phase I and II clinical studies are being carried out using the pan-HDACi panobinostat for the treatment of various hematologic malignancies and solid tumors [158, 159]. The combination of panobinostat with radiation has been shown to induce increased cell apoptosis [160]. There are quite a few clinical trials analyzing the efficacy and safety of such combination therapy. There are ongoing phase I and II clinical trials which include use of radiotherapy with HDACi vorinostat for gastrointestinal tract carcinoma patients; with panobinostat for treatment of prostate, esophageal, and head and neck cancers; with vorinostat against NSCLC; with VPA and temozolomide for high-grade gliomas; and with capeceitabine and vorinostat for treating patients with nonmetastatic pancreatic cancer [154, 161].

Combination therapy with HDACi and DNMTi

Combination therapy involving two major epigenetic modifiers, HDACi and DNMTi is being investigated with great interest. In vitro studies on leukemia cell lines HL-60 and MOLT4 showed that combination of HDACi VPA and DNMTi decitabine induces a synergistic effect with increased cell apoptosis [161]. An ensuing phase I/II clinical trial for the treatment of leukemia patients with VPA and decitabine demonstrated the safety of this combination therapy and provided the platform for pursuing such trials [162]. Another phase I clinical trial for treating refractory solid tumors or hematological malignancies was carried out with HDACi phenylbutyrate and DNMTi 5-Aza which again showed promising results [163]. Likewise, treatment strategy combining HDACi sodium phenylbutyrate with 5-Aza for myelodysplastic syndrome (MDS) or acute myeloid leukemia (AML) patients is also being studied for better patient outcomes [164]. A phase I clinical trial with HDACi VPA in combination with 5-Aza used to treat patients with advanced cancers such as colorectal cancer, melanoma, and breast cancer demonstrated the safety of such combinations in humans [165]. Another phase I clinical trial for MDS or AML patients using HDACi vorinostat in combination with DNMTi decitabine demonstrated promising clinical activity [166]. Further studies are needed to be carried out to determine the efficiency and long-term effects of such combination therapy.

Conclusion

The last three decades have witnessed unparalleled success in identifying a large number of epigenetic changes and their modifiers. The dynamic expression of different genes is regulated through the balance between these epigenetic modifications. The roles of epigenetic modifications have been well characterized and the molecular mechanisms linking these modifications with different physiological processes are being explored. Considerable progress has been made in understanding epigenetic alterations and the epigenetic modifiers related to such modifications in different immunological processes. Role of different epigenetic modifiers involved in different immune functions such as differentiation and plasticity of T cells, cytokine secretion, and other effector functions are well studied. Thus, epigenetic modifiers such as inhibitors of histone deacetylases, DNA methyltransferases, and histone methyltransferases play an important role in immunomodulation. The balance between the different epigenetic modifiers is essential for the tightly regulated functioning of the immune system thereby maintaining the immune homeostasis. Imbalance in the functioning of epigenetic modifiers leads to different immune system-related pathological conditions. Targeting these epigenetic modifiers specifically in cancer has revolutionized the field of cancer treatment. Toxicity of epigenetic drugs, their nonspecific nature, and resistance to these drugs are the main concern of epigenetic modifier based therapy. To overcome these limitations, combinatorial approaches have been adopted that have shown good clinical responses. Preclinical studies on the combination of epigenetic inhibitors with immunotherapeutic agents are well tolerated and show interesting outcomes in cancer patients. To achieve better clinical responses to epigenetic modifier based therapy in different diseases, they need to be further investigated at the molecular level. Further development of novel specific inhibitors which target various epigenetic modifiers with greater efficacy should be explored. Future research on the mechanism of epigenetic modifiers and their role in immune modulation will unravel new targets for epigenetic drugs in order to improve the therapeutic responses and disease-free survival in different pathological conditions.

Acknowledgments

We acknowledge Department of Biotechnology, India and Department of Atomic Energy, Government of India for providing fellowships to Sajad A. Bhat and Shalini K. Sureshbabu. Our work cited in the chapter (Ref. [138]) was supported by the Department of Biotechnology Unit of Excellence, India Grant BT/MED/30/SP11288/2015 to Shubhada Chiplunkar.

Conflict of interest

The authors declare no potential conflicts of interest.

References

[1] Putiri EL, Robertson KD. Epigenetic mechanisms and genome stability. Clin Epigenetics 2011;2(2):299–314.
[2] Jaenisch R, Bird A. Epigenetic regulation of gene expression: how the genome integrates intrinsic and environmental signals. Nat Genet 2003;33(Suppl):245–54.
[3] Skinner MK. Role of epigenetics in developmental biology and transgenerational inheritance. Birth Defects Res C Embryo Today 2011;93(1):51–5.
[4] Kelly TK, De Carvalho DD, Jones PA. Epigenetic modifications as therapeutic targets. Nat Biotechnol 2010;28(10):1069–78.
[5] Strahl BD, Allis CD. The language of covalent histone modifications. Nature 2000;403(6765):41–5.
[6] Portela A, Esteller M. Epigenetic modifications and human disease. Nat Biotechnol 2010;28(10):1057–68.
[7] Khan O, La Thangue NB. HDAC inhibitors in cancer biology: emerging mechanisms and clinical applications. Immunol Cell Biol 2012;90(1):85–94.
[8] True O, Matthias P. Interplay between histone deacetylases and autophagy—from cancer therapy to neurodegeneration. Immunol Cell Biol 2012;90(1):78–84.
[9] Gallagher SJ, Shklovskaya E, Hersey P. Epigenetic modulation in cancer immunotherapy. Curr Opin Pharmacol 2017;35:48–56.
[10] Eckschlager T, et al. Histone deacetylase inhibitors as anticancer drugs. Int J Mol Sci 2017;18(7).
[11] Manzo F, et al. Histone acetyltransferase inhibitors and preclinical studies. Expert Opin Ther Pat 2009;19(6):761–74.
[12] Gros C, et al. DNA methylation inhibitors in cancer: recent and future approaches. Biochimie 2012;94(11):2280–96.
[13] Doroshow DB, Eder JP, LoRusso PM. BET inhibitors: a novel epigenetic approach. Ann Oncol 2017;28(8):1776–87.
[14] French CA, et al. BRD4-NUT fusion oncogene: a novel mechanism in aggressive carcinoma. Cancer Res 2003;63(2):304–7.
[15] Asangani IA, et al. Therapeutic targeting of BET bromodomain proteins in castration-resistant prostate cancer. Nature 2014;510(7504):278–82.

[16] Leal AS, et al. Bromodomain inhibitors, JQ1 and I-BET 762, as potential therapies for pancreatic cancer. Cancer Lett 2017;394:76–87.

[17] Perez-Salvia M, Esteller M. Bromodomain inhibitors and cancer therapy: from structures to applications. Epigenetics 2017;12(5):323–39.

[18] Zhu H, et al. BET bromodomain inhibition promotes anti-tumor immunity by suppressing PD-L1 expression. Cell Rep 2016;16(11):2829–37.

[19] Georgiev P, et al. BET bromodomain inhibition suppresses human T cell function. Immunohorizons 2019;3(7):294–305.

[20] Gan L, et al. Epigenetic regulation of cancer progression by EZH2: from biological insights to therapeutic potential. Biomark Res 2018;6:10.

[21] Gulati N, Beguelin W, Giulino-Roth L. Enhancer of zeste homolog 2 (EZH2) inhibitors. Leuk Lymphoma 2018;59(7):1574–85.

[22] Sun W, et al. Enhancing the anticancer efficacy of immunotherapy through combination with histone modification inhibitors. Genes (Basel) 2018;9(12).

[23] Verde G, et al. Lysine-specific histone demethylases contribute to cellular differentiation and carcinogenesis. Epigenomes 2017;1(1):4.

[24] Liu D, Zempleni J. Low activity of LSD1 elicits a pro-inflammatory gene expression profile in riboflavin-deficient human T Lymphoma Jurkat cells. Genes Nutr 2014;9(5):422.

[25] Yang GJ, et al. Pharmacological inhibition of LSD1 for cancer treatment. Molecules 2018;23(12).

[26] Boulding T, et al. LSD1 activation promotes inducible EMT programs and modulates the tumour microenvironment in breast cancer. Sci Rep 2018;8(1):73.

[27] Konovalov S, Garcia-Bassets I. Analysis of the levels of lysine-specific demethylase 1 (LSD1) mRNA in human ovarian tumors and the effects of chemical LSD1 inhibitors in ovarian cancer cell lines. J Ovarian Res 2013;6(1):75.

[28] Copeland RA, Solomon ME, Richon VM. Protein methyltransferases as a target class for drug discovery. Nat Rev Drug Discov 2009;8(9):724–32.

[29] Nguyen AT, Zhang Y. The diverse functions of Dot1 and H3K79 methylation. Genes Dev 2011;25 (13):1345–58.

[30] Jones B, et al. The histone H3K79 methyltransferase Dot1L is essential for mammalian development and heterochromatin structure. PLoS Genet 2008;4(9):e1000190.

[31] Nguyen AT, et al. DOT1L regulates dystrophin expression and is critical for cardiac function. Genes Dev 2011;25(3):263–74.

[32] Feng Y, et al. Early mammalian erythropoiesis requires the Dot1L methyltransferase. Blood 2010;116 (22):4483–91.

[33] Slany RK. The molecular mechanics of mixed lineage leukemia. Oncogene 2016;35(40):5215–23.

[34] Bernt KM, et al. MLL-rearranged leukemia is dependent on aberrant H3K79 methylation by DOT1L. Cancer Cell 2011;20(1):66–78.

[35] Yao Y, et al. Selective inhibitors of histone methyltransferase DOT1L: design, synthesis, and crystallographic studies. J Am Chem Soc 2011;133(42):16746–9.

[36] Tao R, et al. Deacetylase inhibition promotes the generation and function of regulatory T cells. Nat Med 2007;13(11):1299–307.

[37] Barrett RM, Wood MA. Beyond transcription factors: the role of chromatin modifying enzymes in regulating transcription required for memory. Learn Mem 2008;15(7):460–7.

[38] Hatton RD, et al. A distal conserved sequence element controls Ifng gene expression by T cells and NK cells. Immunity 2006;25(5):717–29.

[39] Roh TY, et al. The genomic landscape of histone modifications in human T cells. Proc Natl Acad Sci U S A 2006;103(43):15782–7.

[40] Akimzhanov AM, Yang XO, Dong C. Chromatin remodeling of interleukin-17 (IL-17)-IL-17F cytokine gene locus during inflammatory helper T cell differentiation. J Biol Chem 2007;282 (9):5969–72.

[41] Wei G, et al. Global mapping of H3K4me3 and H3K27me3 reveals specificity and plasticity in lineage fate determination of differentiating CD4+ T cells. Immunity 2009;30(1):155–67.

[42] Messi M, et al. Memory and flexibility of cytokine gene expression as separable properties of human T(H)1 and T(H)2 lymphocytes. Nat Immunol 2003;4(1):78–86.

[43] Placek K, et al. Genetic and epigenetic networks controlling T helper 1 cell differentiation. Immunology 2009;127(2):155–62.

[44] Ansel KM, et al. Regulation of Th2 differentiation and Il4 locus accessibility. Annu Rev Immunol 2006;24:607–56.

[45] Avni O, et al. T(H) cell differentiation is accompanied by dynamic changes in histone acetylation of cytokine genes. Nat Immunol 2002;3(7):643–51.

[46] Fields PE, Kim ST, Flavell RA. Cutting edge: changes in histone acetylation at the IL-4 and IFN-γ loci accompany Th1/Th2 differentiation. J Immunol 2002;169(2):647–50.

[47] Janson PC, et al. FOXP3 promoter demethylation reveals the committed Treg population in humans. PLoS One 2008;3(2):e1612.

[48] Floess S, et al. Epigenetic control of the foxp3 locus in regulatory T cells. PLoS Biol 2007;5(2):e38.

[49] Polansky JK, et al. DNA methylation controls Foxp3 gene expression. Eur J Immunol 2008;38(6):1654–63.

[50] Hassig CA, et al. A role for histone deacetylase activity in HDAC1-mediated transcriptional repression. Proc Natl Acad Sci U S A 1998;95(7):3519–24.

[51] Ashburner BP, Westerheide SD, Baldwin Jr AS. The p65 (RelA) subunit of NF-kappaB interacts with the histone deacetylase (HDAC) corepressors HDAC1 and HDAC2 to negatively regulate gene expression. Mol Cell Biol 2001;21(20):7065–77.

[52] Zhong H, et al. The phosphorylation status of nuclear NF-kappa B determines its association with CBP/p300 or HDAC-1. Mol Cell 2002;9(3):625–36.

[53] Li Q, Verma IM. NF-kappaB regulation in the immune system. Nat Rev Immunol 2002;2(10):725–34.

[54] Powell JD, et al. The -180 site of the IL-2 promoter is the target of CREB/CREM binding in T cell anergy. J Immunol 1999;163(12):6631–9.

[55] Telander DG, Malvey EN, Mueller DL. Evidence for repression of IL-2 gene activation in anergic T cells. J Immunol 1999;162(3):1460–5.

[56] Moreira JM, Scheipers P, Sorensen P. The histone deacetylase inhibitor Trichostatin A modulates CD4+ T cell responses. BMC Cancer 2003;3:30.

[57] Skov S, et al. Histone deacetylase inhibitors: a new class of immunosuppressors targeting a novel signal pathway essential for CD154 expression. Blood 2003;101(4):1430–8.

[58] Dangond F, Gullans SR. Differential expression of human histone deacetylase mRNAs in response to immune cell apoptosis induction by trichostatin A and butyrate. Biochem Biophys Res Commun 1998;247(3):833–7.

[59] Grausenburger R, et al. Conditional deletion of histone deacetylase 1 in T cells leads to enhanced airway inflammation and increased Th2 cytokine production. J Immunol 2010;185(6):3489–97.

[60] Glauben R, et al. Histone deacetylase inhibitors modulate interleukin 6-dependent CD4+ T cell polarization in vitro and in vivo. J Biol Chem 2014;289(9):6142–51.

[61] Bhat J, et al. Histone deacetylase inhibitor modulates NKG2D receptor expression and memory phenotype of human gamma/delta T cells upon interaction with tumor cells. Front Immunol 2019;10:569.

[62] Bhat J, Oberg HH, Kabelitz D. Modulation of human gamma/delta T-cell activation and phenotype by histone deacetylase inhibitors. Cell Immunol 2015;296(1):50–6.

[63] Bhat J, et al. Immunotherapy with human gamma delta T cells-synergistic potential of epigenetic drugs? Front Immunol 2018;9:512.

[64] de Zoeten EF, Wang L, Butler K, et al. Histone deacetylase 6 and heat shock protein 90 control the functions of Foxp3(+) T-regulatory cells. Mol Cell Biol 2011;31(10):2066–78. https://doi.org/10.1128/MCB.05155-11.

[65] Li B, Greene MI. FOXP3 actively represses transcription by recruiting the HAT/HDAC complex. Cell Cycle 2007;6(12):1432–6.

[66] Shen L, et al. Class I histone deacetylase inhibitor entinostat suppresses regulatory T cells and enhances immunotherapies in renal and prostate cancer models. PLoS One 2012;7(1):e30815.

[67] Moon C, et al. Use of epigenetic modification to induce FOXP3 expression in naive T cells. Transplant Proc 2009;41(5):1848–54.

[68] Magner WJ, et al. Activation of MHC class I, II, and CD40 gene expression by histone deacetylase inhibitors. J Immunol 2000;165(12):7017–24.

[69] Orillion A, et al. Entinostat neutralizes myeloid-derived suppressor cells and enhances the antitumor effect of PD-1 inhibition in murine models of lung and renal cell carcinoma. Clin Cancer Res 2017;23(17):5187–201.

[70] Brogdon JL, et al. Histone deacetylase activities are required for innate immune cell control of Th1 but not Th2 effector cell function. Blood 2007;109(3):1123–30.

[71] Cheng F, et al. A novel role for histone deacetylase 6 in the regulation of the tolerogenic STAT3/IL-10 pathway in APCs. J Immunol 2014;193(6):2850–62.

[72] Cheng F, et al. Divergent roles of histone deacetylase 6 (HDAC6) and histone deacetylase 11 (HDAC11) on the transcriptional regulation of IL10 in antigen presenting cells. Mol Immunol 2014;60(1):44–53.

[73] Akimova T, et al. Histone/protein deacetylases and T-cell immune responses. Blood 2012;119(11):2443–51.

[74] Mullican SE, et al. Histone deacetylase 3 is an epigenomic brake in macrophage alternative activation. Genes Dev 2011;25(23):2480–8.

[75] Iglesias MJ, et al. Combined chromatin and expression analysis reveals specific regulatory mechanisms within cytokine genes in the macrophage early immune response. PLoS One 2012;7(2):e32306.

[76] Foster SL, Hargreaves DC, Medzhitov R. Gene-specific control of inflammation by TLR-induced chromatin modifications. Nature 2007;447(7147):972–8.

[77] Schmidt SV, et al. The transcriptional regulator network of human inflammatory macrophages is defined by open chromatin. Cell Res 2016;26(2):151–70.

[78] Das Gupta K, et al. Histone deacetylases in monocyte/macrophage development, activation and metabolism: refining HDAC targets for inflammatory and infectious diseases. Clin Transl Immunol 2016;5(1):e62.

[79] Nencioni A, et al. Histone deacetylase inhibitors affect dendritic cell differentiation and immunogenicity. Clin Cancer Res 2007;13(13):3933–41.

[80] Wang B, et al. Butyrate inhibits functional differentiation of human monocyte-derived dendritic cells. Cell Immunol 2008;253(1-2):54–8.

[81] Bosisio D, et al. Blocking TH17-polarizing cytokines by histone deacetylase inhibitors in vitro and in vivo. J Leukoc Biol 2008;84(6):1540–8.

[82] Jung ID, et al. Apicidin, the histone deacetylase inhibitor, suppresses Th1 polarization of murine bone marrow-derived dendritic cells. Int J Immunopathol Pharmacol 2009;22(2):501–15.

[83] Stubig T, et al. 5-azacytidine promotes an inhibitory T-cell phenotype and impairs immune mediated antileukemic activity. Mediat Inflamm 2014;2014:418292.

[84] Wilson CB, Rowell E, Sekimata M. Epigenetic control of T-helper-cell differentiation. Nat Rev Immunol 2009;9(2):91–105.

[85] Lee PP, et al. A critical role for Dnmt1 and DNA methylation in T cell development, function, and survival. Immunity 2001;15(5):763–74.

[86] Thomas RM, et al. De novo DNA methylation is required to restrict T helper lineage plasticity. J Biol Chem 2012;287(27):22900–9.

[87] Chappell C, et al. DNA methylation by DNA methyltransferase 1 is critical for effector CD8 T cell expansion. J Immunol 2006;176(8):4562–72.

[88] Emran AA, et al. Targeting DNA methylation and EZH2 activity to overcome melanoma resistance to immunotherapy. Trends Immunol 2019;40(4):328–44.

[89] Costantini B, et al. The effects of 5-azacytidine on the function and number of regulatory T cells and T-effectors in myelodysplastic syndrome. Haematologica 2013;98(8):1196–205.

[90] Barwick BG, et al. B cell activation and plasma cell differentiation are inhibited by de novo DNA methylation. Nat Commun 2018;9(1):1900.

[91] Wu H, et al. Epigenetic regulation in B-cell maturation and its dysregulation in autoimmunity. Cell Mol Immunol 2018;15(7):676–84.

[92] Zheng X, et al. Redirecting tumor-associated macrophages to become tumoricidal effectors as a novel strategy for cancer therapy. Oncotarget 2017;8(29):48436–52.

[93] Yang X, et al. Epigenetic regulation of macrophage polarization by DNA methyltransferase 3b. Mol Endocrinol 2014;28(4):565–74.

[94] de Groot AE, Pienta KJ. Epigenetic control of macrophage polarization: implications for targeting tumor-associated macrophages. Oncotarget 2018;9(29):20908–27.

[95] Jeong HY, et al. 5-Azacytidine modulates interferon regulatory factor 1 in macrophages to exert a cardioprotective effect. Sci Rep 2015;5:15768.

[96] Kim YS, et al. Protective role of 5-azacytidine on myocardial infarction is associated with modulation of macrophage phenotype and inhibition of fibrosis. J Cell Mol Med 2014;18(6):1018–27.

[97] Kim HD, et al. Heat shock and 5-azacytidine inhibit nitric oxide synthesis and tumor necrosis factor-alpha secretion in activated macrophages. Antioxid Redox Signal 1999;1(3):297–304.

[98] Cao Q, et al. Inhibiting DNA Methylation by 5-Aza-2'-deoxycytidine ameliorates atherosclerosis through suppressing macrophage inflammation. Endocrinology 2014;155(12):4925–38.

[99] Thangavel J, et al. Epigenetic modifiers reduce inflammation and modulate macrophage phenotype during endotoxemia-induced acute lung injury. J Cell Sci 2015;128(16):3094–105.

[100] Frikeche J, et al. Impact of the hypomethylating agent 5-azacytidine on dendritic cells function. Exp Hematol 2011;39(11):1056–63.

[101] Busslinger M, Tarakhovsky A. Epigenetic control of immunity. Cold Spring Harb Perspect Biol 2014;6(6).

[102] Maeda T, et al. Up-regulation of costimulatory/adhesion molecules by histone deacetylase inhibitors in acute myeloid leukemia cells. Blood 2000;96(12):3847–56.

[103] Skov S, et al. Cancer cells become susceptible to natural killer cell killing after exposure to histone deacetylase inhibitors due to glycogen synthase kinase-3-dependent expression of MHC class I-related chain A and B. Cancer Res 2005;65(23):11136–45.

[104] Armeanu S, et al. Natural killer cell-mediated lysis of hepatoma cells via specific induction of NKG2D ligands by the histone deacetylase inhibitor sodium valproate. Cancer Res 2005;65 (14):6321–9.

[105] Lopez-Soto A, et al. HDAC3 represses the expression of NKG2D ligands ULBPs in epithelial tumour cells: potential implications for the immunosurveillance of cancer. Oncogene 2009;28(25):2370–82.

[106] Buglio D, et al. HDAC11 plays an essential role in regulating OX40 ligand expression in Hodgkin lymphoma. Blood 2011;117(10):2910–7.

[107] Karpf AR. A potential role for epigenetic modulatory drugs in the enhancement of cancer/germ-line antigen vaccine efficacy. Epigenetics 2006;1(3):116–20.

[108] Khan AN, Magner WJ, Tomasi TB. An epigenetic vaccine model active in the prevention and treatment of melanoma. J Transl Med 2007;5:64.

[109] Pellicciotta I, et al. Presentation of telomerase reverse transcriptase, a self-tumor antigen, is down-regulated by histone deacetylase inhibition. Cancer Res 2008;68(19):8085–93.

[110] de Zoeten EF, et al. Inhibition of HDAC9 increases T regulatory cell function and prevents colitis in mice. Gastroenterology 2010;138(2):583–94.

[111] Glauben R, et al. Histone hyperacetylation is associated with amelioration of experimental colitis in mice. J Immunol 2006;176(8):5015–22.

[112] Choi SW, et al. Histone deacetylase inhibition regulates inflammation and enhances Tregs after allogeneic hematopoietic cell transplantation in humans. Blood 2015;125(5):815–9.

[113] Aspeslagh S, et al. Epigenetic modifiers as new immunomodulatory therapies in solid tumours. Ann Oncol 2018;29(4):812–24.

[114] Cihak A. Biological effects of 5-azacytidine in eukaryotes. Oncology 1974;30(5):405–22.

[115] Kantarjian HM, et al. Results of decitabine (5-aza-2'deoxycytidine) therapy in 130 patients with chronic myelogenous leukemia. Cancer 2003;98(3):522–8.

[116] Winquist E, et al. Phase II trial of DNA methyltransferase 1 inhibition with the antisense oligonucleotide MG98 in patients with metastatic renal carcinoma: a National Cancer Institute of Canada Clinical Trials Group investigational new drug study. Investig New Drugs 2006;24(2):159–67.

[117] Blum KA, et al. Phase I trial of low dose decitabine targeting DNA hypermethylation in patients with chronic lymphocytic leukaemia and non-Hodgkin lymphoma: dose-limiting myelosuppression without evidence of DNA hypomethylation. Br J Haematol 2010;150(2):189–95.

[118] Clozel T, et al. Mechanism-based epigenetic chemosensitization therapy of diffuse large B-cell lymphoma. Cancer Discov 2013;3(9):1002–19.

[119] Mathilde C, et al. DNMT inhibitors in cancer, current treatments and future promising approach: inhibition of specific DNMT-including complexes. Epigenet Diagn Ther 2015;1(1):37–48.

[120] Hubaux R, et al. EZH2 promotes E2F-driven SCLC tumorigenesis through modulation of apoptosis and cell-cycle regulation. J Thorac Oncol 2013;8(8):1102–6.
[121] Shen X, et al. Jumonji modulates polycomb activity and self-renewal versus differentiation of stem cells. Cell 2009;139(7):1303–14.
[122] Lee TI, et al. Control of developmental regulators by Polycomb in human embryonic stem cells. Cell 2006;125(2):301–13.
[123] Varambally S, et al. Genomic loss of microRNA-101 leads to overexpression of histone methyltransferase EZH2 in cancer. Science 2008;322(5908):1695–9.
[124] Simon JA, Lange CA. Roles of the EZH2 histone methyltransferase in cancer epigenetics. Mutat Res 2008;647(1-2):21–9.
[125] Daigle SR, et al. Selective killing of mixed lineage leukemia cells by a potent small-molecule DOT1L inhibitor. Cancer Cell 2011;20(1):53–65.
[126] Morera L, Lubbert M, Jung M. Targeting histone methyltransferases and demethylases in clinical trials for cancer therapy. Clin Epigenetics 2016;8:57.
[127] Daigle SR, et al. Potent inhibition of DOT1L as treatment of MLL-fusion leukemia. Blood 2013;122(6):1017–25.
[128] Klaus CR, et al. DOT1L inhibitor EPZ-5676 displays synergistic antiproliferative activity in combination with standard of care drugs and hypomethylating agents in MLL-rearranged leukemia cells. J Pharmacol Exp Ther 2014;350(3):646–56.
[129] Basavapathruni A, et al. Nonclinical pharmacokinetics and metabolism of EPZ-5676, a novel DOT1L histone methyltransferase inhibitor. Biopharm Drug Dispos 2014;35(4):237–52.
[130] Kartikasari AER, et al. Therapeutic cancer vaccines-T cell responses and epigenetic modulation. Front Immunol 2018;9:3109.
[131] Gameiro SR, et al. Inhibitors of histone deacetylase 1 reverse the immune evasion phenotype to enhance T-cell mediated lysis of prostate and breast carcinoma cells. Oncotarget 2016;7(7):7390–402.
[132] McCaw TR, et al. Modulation of antitumor immunity with histone deacetylase inhibitors. Immunotherapy 2017;9(16):1359–72.
[133] Terranova-Barberio M, Thomas S, Munster PN. Epigenetic modifiers in immunotherapy: a focus on checkpoint inhibitors. Immunotherapy 2016;8(6):705–19.
[134] Mazzone R, et al. Epi-drugs in combination with immunotherapy: a new avenue to improve anticancer efficacy. Clin Epigenetics 2017;9:59.
[135] Chiappinelli KB, et al. Combining epigenetic and immunotherapy to combat cancer. Cancer Res 2016;76(7):1683–9.
[136] Hammers HJ, et al. Safety and efficacy of nivolumab in combination with ipilimumab in metastatic renal cell carcinoma: the CheckMate 016 study. J Clin Oncol 2017;35(34):3851–8.
[137] Larkin J, et al. Combined nivolumab and ipilimumab or monotherapy in untreated melanoma. N Engl J Med 2015;373(1):23–34.
[138] Bhat SA, Vedpathak DM, Chiplunkar SV. Checkpoint blockade rescues the repressive effect of histone deacetylases inhibitors on gammadelta T cell function. Front Immunol 2018;9:1615.
[139] Kim K, et al. Eradication of metastatic mouse cancers resistant to immune checkpoint blockade by suppression of myeloid-derived cells. Proc Natl Acad Sci U S A 2014;111(32):11774–9.
[140] Lisiero DN, et al. The histone deacetylase inhibitor, LBH589, promotes the systemic cytokine and effector responses of adoptively transferred CD8+ T cells. J Immunother Cancer 2014;2:8.
[141] Li H, et al. Immune regulation by low doses of the DNA methyltransferase inhibitor 5-azacitidine in common human epithelial cancers. Oncotarget 2014;5(3):587–98.
[142] Yang H, et al. Expression of PD-L1, PD-L2, PD-1 and CTLA4 in myelodysplastic syndromes is enhanced by treatment with hypomethylating agents. Leukemia 2014;28(6):1280–8.
[143] Catalano MG, et al. Valproic acid enhances tubulin acetylation and apoptotic activity of paclitaxel on anaplastic thyroid cancer cell lines. Endocr Relat Cancer 2007;14(3):839–45.
[144] Dowdy SC, et al. Histone deacetylase inhibitors and paclitaxel cause synergistic effects on apoptosis and microtubule stabilization in papillary serous endometrial cancer cells. Mol Cancer Ther 2006;5(11):2767–76.
[145] Ramalingam SS, et al. Phase I and pharmacokinetic study of vorinostat, a histone deacetylase inhibitor, in combination with carboplatin and paclitaxel for advanced solid malignancies. Clin Cancer Res 2007;13(12):3605–10.

[146] Pauer LR, et al. Phase I study of oral CI-994 in combination with carboplatin and paclitaxel in the treatment of patients with advanced solid tumors. Cancer Investig 2004;22(6):886–96.

[147] Ramalingam SS, et al. Carboplatin and Paclitaxel in combination with either vorinostat or placebo for first-line therapy of advanced non-small-cell lung cancer. J Clin Oncol 2010;28(1):56–62.

[148] Yardley DA, et al. Randomized phase II, double-blind, placebo-controlled study of exemestane with or without entinostat in postmenopausal women with locally recurrent or metastatic estrogen receptor-positive breast cancer progressing on treatment with a nonsteroidal aromatase inhibitor. J Clin Oncol 2013;31(17):2128–35.

[149] Lee JH, et al. Histone deacetylase inhibitor enhances 5-fluorouracil cytotoxicity by down-regulating thymidylate synthase in human cancer cells. Mol Cancer Ther 2006;5(12):3085–95.

[150] Sung MW, Waxman S. Combination of cytotoxic-differentiation therapy with 5-fluorouracil and phenylbutyrate in patients with advanced colorectal cancer. Anticancer Res 2007;27(2):995–1001.

[151] Undevia SD, et al. A phase I study of the oral combination of CI-994, a putative histone deacetylase inhibitor, and capecitabine. Ann Oncol 2004;15(11):1705–11.

[152] Thurn KT, et al. Rational therapeutic combinations with histone deacetylase inhibitors for the treatment of cancer. Future Oncol 2011;7(2):263–83.

[153] Chinnaiyan P, et al. Modulation of radiation response by histone deacetylase inhibition. Int J Radiat Oncol Biol Phys 2005;62(1):223–9.

[154] Groselj B, et al. Histone deacetylase inhibitors as radiosensitisers: effects on DNA damage signalling and repair. Br J Cancer 2013;108(4):748–54.

[155] Arundel CM, Glicksman AS, Leith JT. Enhancement of radiation injury in human colon tumor cells by the maturational agent sodium butyrate (NaB). Radiat Res 1985;104(3):443–8.

[156] Shabason JE, Tofilon PJ, Camphausen K. Grand rounds at the National Institutes of Health: HDAC inhibitors as radiation modifiers, from bench to clinic. J Cell Mol Med 2011;15(12):2735–44.

[157] Noguchi H, et al. Successful treatment of anaplastic thyroid carcinoma with a combination of oral valproic acid, chemotherapy, radiation and surgery. Endocr J 2009;56(2):245–9.

[158] Venturelli S, et al. Epigenetic combination therapy as a tumor-selective treatment approach for hepatocellular carcinoma. Cancer 2007;109(10):2132–41.

[159] Prince HM, Bishton MJ, Johnstone RW. Panobinostat (LBH589): a potent pan-deacetylase inhibitor with promising activity against hematologic and solid tumors. Future Oncol 2009;5(5):601–12.

[160] Xiao W, et al. Combination therapy with the histone deacetylase inhibitor LBH589 and radiation is an effective regimen for prostate cancer cells. PLoS One 2013;8(8):e74253.

[161] Suraweera A, O'Byrne KJ, Richard DJ. Combination therapy with histone deacetylase inhibitors (HDACi) for the treatment of cancer: achieving the full therapeutic potential of HDACi. Front Oncol 2018;8:92.

[162] Garcia-Manero G, et al. Phase 1/2 study of the combination of 5-aza-2'-deoxycytidine with valproic acid in patients with leukemia. Blood 2006;108(10):3271–9.

[163] Rudek MA, et al. Pharmacokinetics of 5-azacitidine administered with phenylbutyrate in patients with refractory solid tumors or hematologic malignancies. J Clin Oncol 2005;23(17):3906–11.

[164] Gore SD, et al. Combined DNA methyltransferase and histone deacetylase inhibition in the treatment of myeloid neoplasms. Cancer Res 2006;66(12):6361–9.

[165] Braiteh F, et al. Phase I study of epigenetic modulation with 5-azacytidine and valproic acid in patients with advanced cancers. Clin Cancer Res 2008;14(19):6296–301.

[166] Kirschbaum M, et al. A phase 1 clinical trial of vorinostat in combination with decitabine in patients with acute myeloid leukaemia or myelodysplastic syndrome. Br J Haematol 2014;167(2):185–93.

Index

Note: Page numbers followed by *f* indicate figures, and *t* indicate tables.

A

Accessible chromatin, epigenetic modifications, 44
Acinetobacter lwoffii F78, 152
Acute myeloid leukemia (AML), 205–206
  chromosomal aberrations, 294
  complex aberrant karyotype AML (ck-AML), 296
  definition, 293
  deletions/monosomies, 295
  DNA methylation profiles, 294
  enhancer hijacking, 296
  genetic and genomic aberrations, 294
  genomic amplifications, 296
  mixed-lineage leukemia (MLL), 294–295
  radio-/chemotherapy, 293
Adaptive immune system, 1–2, 51–63
  B-cell development and differentiation, 52–55, 53*f*
  memory in, 9*f*
  T cells, epigenetic regulation in, 55–60
AgIn, 175
Aging, single-cell epigenomics technologies, 206–208
Altered Schaedler flora (ASF), 142–144
Alzheimer's disease (AD), 207
Alzheimer's disease-associated microglia (DAM), 207
Anchor sites, 87
Anti-cyclic citrullinated peptide (anti-CCP), 254–255
Antimicrobial pathogenesis-related proteins (PR), 29–30
Anti-TNF therapy, 255
Arabidopsis thaliana L., 26–28, 31*t*, 34–35
Area under the ROC curve (AUC), 306–307
Arginine and lysinemethyltransferases, 18
Argonaute (AGO) protein, 90
Artificial intelligence (AI), 218
Assay for transposase-accessible chromatin (ATAC-seq), 11, 44–45, 57–58, 63, 194–195, 206–207
Atezolizumab, 318, 329–340*t*
Autoimmune diseases, epigenetic deregulation in, 254–258
  in psoriasis, 256–258
  rheumatoid arthritis (RA), 254–256
Autoimmune regulator (AIRE), 2–3
Autoimmunity, epigenetic biomarkers in, 260–262, 261*t*
Autoinflammatory diseases, epigenetic deregulation in, 258–260
  cryopyrin-associated periodic syndromes (CAPS), 259–260
  Familial Mediterranean fever (FMF), 258–259
Avelumab, 329–340*t*
Azacitidine, 329–340*t*

B

Bacillus Calmette Guérin (BCG) vaccination, 8–9
Bacteriocins, 148
Bacteroidetes, 142–144
Basal/horizontal disease resistance, 32–33
Base excision repair (BER) pathway, 28–29
Bayesian networks, 227–229
B-cell development, 3–4, 52–55
  common lymphoid progenitor cells (CLP), 52
  Foxo1, 53–54
  PAX5, 52–53
  peripheral differentiation, 54–55
  pre-B to immature B cells, 54
  pro-B to pre-B cell commitment, 54
  transcriptional and epigenetic regulation of, 53*f*
B-cell differentiation, 52–55
  common lymphoid progenitor cells (CLP), 52
  Foxo1, 53–54
  long noncoding RNAs (lncRNA), 103–104
  microRNA (miRNAs)-mediated regulation of, 90–91
  PAX5, 52–53
  peripheral differentiation, 54–55
  pre-B to immature B cells, 54
  pro-B to pre-B cell commitment, 54
  specific microRNA (miRNAs) regulating, 92–100
  transcriptional and epigenetic regulation of, 53*f*
B-cell integration cluster (BIC), 105
B-cell receptor (BCR), 3–4, 54, 78
Belinostat, 327–328*t*

Bendamustine, 329–340*t*
β-glucan, 145–146
β-score, 175
BIGpre, 173
Binary Alignment/Map (BAM), 174
Biotic stresses
  epigenetics of
    immune system and memory in plants, 32–33
    plant insect interactions, 30–32, 31*t*
    plant microbe interactions, 29–30
  histone modifications, 29
  plant DNA methylation, 26–27, 27*f*
    de novo DNA methylation, 27–28
    maintenance, 28–29
  plant epigenetics, 33
    heterosis, 34–35
    somatic embryogenesis, 34
  on plants influence epigenetic mechanisms, 25, 26*f*
  RNA-associated silencing, 29
Bisulfite sequencing read aligners, 174
Bisulfite-treated Reads Analysis Tool (BRAT), 174
Bivalent *Cebpa* promoter, 57
Bivalent epigenetic mark, 44–45
Bivalent genes, 85
Bivalent loci, 121
Bivalent promoters, 48–49
Blimp-1, 54–55
Bortezomib, 329–340*t*
Bowtie 2, 198
BRAT using the Burrows-Wheeler transform (BRAT-BW), 174
Bromodomain and extra-terminal (BET) inhibitors, 317–318
Bromodomain and extraterminal (BET) proteins, 14–15, 87
Bromodomain (Brd) proteins, 14–15
BSeQC, 173
Butyrate, 142–144

C

Candida albicans, 145–146
Canonical histone, 16–17
Carboplatin, 329–340*t*
CCAAT enhancer binding protein (C/EBP), 48–49
CCAAT/enhancer binding protein alpha (CBPa), 51
CCCTC-binding factor (CTCF), 127–128
CD62L (L-selectin), 119–120
CD4+ T cells, 6, 8, 55, 58–59, 78–79, 91, 117, 124, 142–144
CD8+ T cells, 6–7, 55, 59, 82–83, 88–89, 91, 93–98*t*, 117, 123
  effector *vs.* memory fate decisions, 129–132, 131*f*
  function, acquisition of, 124–126, 125*f*
Cebpa promoter, 48–49
Central memory T cells (T_{CM}), 119–120
Central tolerance, 2–3
CG methylation, 28, 30–31
Chemotherapeutics, epigenetic modifiers on immune system, 343–344
Chidamide, 327–328*t*
Childhood asthma. *See* DNA methylation (DNAm)
ChIP-seq. *See* Chromatin immunoprecipitation followed by sequencing (ChIP-seq)
Chromatin, 11
  accessibility, 88–89, 141
  active regulation, in CD8+ T-cell effector *vs.* memory fate decisions, 129–132, 131*f*
  defined, 120
  structure, single-cell epigenomics, 189–190*t*, 193–196
    chromosome conformation, 195
    single-cell multiomics, 195–196
Chromatin-based epigenetic mechanisms in lymphocyte differentiation, 80–89
  chromatin accessibility in immune cells differentiation, 88–89
  DNA methylation and lineage commitment, 80–81, 80*f*
  histone modifications, 84–86
Chromatin folding, noncoding regulatory elements work via, 122–123
Chromatin immunoprecipitation (ChIP), 171
Chromatin immunoprecipitation-exonuclease (ChIP-exo), 171
Chromatin immunoprecipitation followed by sequencing (ChIP-seq), 11, 63, 129, 171, 173–174, 193
  data analysis, 176–177
  quality control of, 177
Chromatin interaction analysis by paired-end tag sequencing (ChIA-PET), 171
Chromatin looping in Th2 cytokine, 62
Chromatin loops, 13–14
Chromatin packaging, Trans-regulation, 11–13, 12*f*
Chromatin remodeling, 16–18
  CD8+ T-cell-specific transcription factors, 127–129
  epigenetic modifications, 44

Chromatosome, 11
Chromomethylase 2 (CMT2), 28
Chromomethylase 3 (CMT3), 28
Chromosomal territories (CTs), 13–14, 61*f*
Chromosome conformation, 189–190*t*, 195
Chromosome Conformation Capture (3C), 195
chromVAR, 200–202
Chronic lymphocytic leukemia (CLL), 205–206, 341
  cancer-specific DNA methylation, 303, 303*f*
  deregulated genes, 301–302
  malignant transformation, 300–301
  methylomics, 302–303
Cis-regulation through DNA elements, 10–11
CITE-seq, 63
Cladribine, 329–340*t*
Class switching, 3–4
Class switch recombination (CSR), 78
Cohesin, 13–14
Collagen-induced arthritis (CIA), 87
Colonic lamina propria macrophages, 144–145
Colorectal neoplasia differentially expressed (CRNDE), 103
Commensals, 148, 149*f*
Common dendritic cell progenitor (CDP), 51
Common lymphoid progenitor (CLP) cells, 2, 3*f*, 4–5, 46–47, 52, 78, 79*f*
  B cells development, 52
  natural killer (NK) cells, development of, 46–47, 46*f*
  to T cells, 55–57, 56*f*
Common myeloid progenitor (CMP) cells, 2, 3*f*
Complex aberrant karyotype AML (ck-AML), 296
Conserved noncoding sequence (CNS), 47
Contact domains, 13–14
Conventional dendritic cells (cDCs), 51
Convolutional neural networks (cNN), 231–232
Covalent modifications, 39
CpG island methylator phenotype (CIMP-positive), 298
CpG site. *See* Cytosines adjacent to guanines (CpG sites)
CRISPR-Cas9-based epigenomic regulatory element screening (CERES), 209
CRM Activity Database 2 (CAD2), 227
Cryopyrin, 259–260
Cryopyrin-associated periodic syndromes (CAPS), 259–260
Cutting-edge technologies, 19–20, 19*f*
CXCL6 locus, 147–148
Cyclin-dependent kinase inhibitor p21 (Cdkn1a) gene, 144
Cyclophosphamide, 329–340*t*
Cytarabine, 329–340*t*, 342
Cytokine-producing cells, 4
Cytosine methylation, 26–28
Cytosines adjacent to guanines (CpG sites), 26–27, 253–254, 260–261
Cytotoxic lymphocytes, classes of, 7*f*

D

Danger-associated molecular pattern molecules (DAMPs), 259–260
Daunorubicin, 342
Decision systems, 219
Decitabine, 329–340*t*, 341–343
Demeter (DME), 28–29
Dendritic cells (DCs), 5–6, 142–145, 146*f*, 240–241
  conventional DCs (cDCs), 51
  plasmacytoid DCs (pDCs), 51
  transcription factors and epigenetic mechanisms in, 51
De novo DNA methylation, 27–28, 82–83
DESeq2, 177–178
Dicer, 54, 90–91
Dicer-like 3 (DCL3), 29
Differentially methylated CpGs (DMCs), 274–275
Differentially methylated cytosines (DMCs), 174–176
Differentially methylated regions (DMRs), 174–176, 230, 273
Differentiation process, 1–2
Diffuse large B-cell lymphoma (DLBCL), 341
Disruptor of telomeric silencing 1-like (DOT1L) inhibitors, 319
Distal regulatory elements, 10–11
DNA adenine methyltransferase identification sequencing (DamID-seq), 193
DNA demethylation, 47, 80–81
DNA hydroxymethylation. *See* DNA methylation (DNAm)
DNA hypomethylation, 207
DNA methylation (DNAm), 11, 25–26, 54, 59, 141, 147–148, 151, 161–162, 253–257, 259–260
  CD8 T cells, 241
  childhood allergic disease and asthma
    airways, 278–279, 280*f*
    allergic sensitization and poly-sensitization, 277–278

DNA methylation (DNAm) *(Continued)*
    differentially methylated CpGs (DMCs), 274–275
    differentially methylated regions (DMRs), 273
    DNA methylation age (DNAmAge), 275–276
    environmental exposures, 276–277
    eosinophils, 277
    epigenome-wide studies, 271–272*t*
    gene expression, 277
    immunoregulatory *(SMAD3)* and proinflammatory *(IL1B)* pathways, 273–274
    Inner-City Asthma Consortium (ICAC) cohort, 274
    lung function parameters, 274
    multi-cohort studies, 279–281
    perinatal and postnatal factors, 270
    technical issues, 281
  epigenetic modifications, 41–42
  junk DNA, making sense of, 122–123
  and lineage commitment, 80–81, 80*f*
  macrophage differentiation and functions, 240–241
  of *MEFV* gene, 258–259
  methylated CpG, 239–240
  methylation data analysis
    differential methylation, 175–176
    methylome segmentation, 176
    scoring, 175
  myeloid lineages, 240
  peripheral blood mononuclear cells (PBMCs), 255
  in regulating gene transcription, 122–123
  self-tolerance and tissue homeostasis, 241
  sequencing technologies for epigenetics, 170
  tumor microenvironment, 241–242
  writers, readers, and erasers in immune cell differentiation, 82–84
DNA methylation age (DNAmAge), 275–276
DNA methylation inhibitor, 145
DNA methyltransferase 3b (DNMT3b), 240–241
DNA methyltransferases (DNMTs), 18, 80–82, 122, 253–254
DNA methyltransferase inhibitors ((DNMTi), 316–317, 324–325
  B cells, 324
  clinical significance of, 341
  combination therapy, and HDACi, 344–345
  dendritic cells (DCs), 325, 326*f*
  immunomodulatory effects, 324–325
  macrophages, 325
  regulatory T (Treg) cells, 324
  T cells, 324
DNA modifications, 189–190*t*
  single-cell epigenomics technologies, 192–193
DNAse Hypersensitive Site (DHS), 250
DNase hypersensitivity assays, 84–85, 172
DNaseI hypersensitivity assay, 145–146
DNase-seq, 194
Domains Rearranged Methyltransferase 2 (DRM2), 28
Dominance heterosis, 34
Doxorubicin, 344
Droplet-based microfluidics protocols, 191
Droplet single-cell assay for transposase-accessible chromatin using sequencing (dsc(i) ATAC-seq), 203–204
Duplicate reads, PCR, 165
Durvalumab, 329–340*t*
Dysbiosis, 139–141, 148

E

Effector memory T cells (T_{EM}), 120
Effector-triggered immunity (ETI), 32–33
Embryonic stem cells (ESCs), 121
Enhancer hijacking, 296
Enhancer of Estes homolog-2 (EZH2), 129
Enhancer regions, 10–11
Enhancer RNAs (eRNAs), 226
Enhancers, 14–15
Entinostat, 329–340*t*
Environmental memories, 32–33
Enzymatic-Methyl-sequencing (EM-seq) technique, 192–193
Eomesodermin (Eomes), 92
Epi-drugs, 262–263
Epigenetic biomarkers, in autoimmunity, 260–262, 261*t*
Epigenetic code, 17
Epigenetic deregulation, 253–254
  in autoimmune diseases, 254–258
    in psoriasis, 256–258
    rheumatoid arthritis (RA), 254–256
  in autoinflammatory diseases, 258–260
    cryopyrin-associated periodic syndromes (CAPS), 259–260
    Familial Mediterranean fever (FMF), 258–259
  epigenetic biomarkers in autoimmunity, 260–262, 261*t*
  targeting epigenetic defects, 262–263, 262*t*

Epigenetic landscape, 10
Epigenetic landscape profiling using cytometry by time-of-flight (EpiTOF), 202
Epigenetic mapping, T-cell memory, 126–127
Epigenetic modifications
  accessible chromatin, 44
  chromatin remodeling and noncoding RNA (ncRNA), 44
  DNA methylation, 41–42
  histone modifications, 42–43
    interdependence of DNA and, 43
    key epigenetic drivers, 42–43
  priming gene for activation, 44–45
  transcriptional control of, 41
Epigenetic modifiers on immune system, 315–319
  bromodomain and extra-terminal (BET) inhibitors, 317–318
  clinical significance of, 325–342, 327–340*t*
    DNA methyltransferase (DNMT) inhibitors therapy, 341
    histone deacetylases (HDAC) inhibitors therapy, 325–341
  disruptor of telomeric silencing 1-like (DOT1L) inhibitors, 319
  DNA methyltransferase (DNMT) inhibitors, 317
  EZH2 inhibitors, 318
  histone acetylase transferase (HAT) inhibitors, 317
  histone deacetylases (HDAC) inhibitors, 317
  immunomodulatory effects, 319–325
    DNA methyltransferase (DNMT) inhibitors on immune cells, 324–325
    HDAC inhibitors (HDACi) on immune cells, 321–323
  lysine-specific demethylase 1 (LSD1) inhibitors, 319
Epigenetic recombinant inbred lines (epi-RILs), 34–35
Epigenetic regulation, 9–10, 45–63
  CD8 T-cell function, acquisition of, 124–126, 125*f*
  hematopoiesis, epigenetic regulation, 45
  innate immune cells, development of, 45–51
    dendritic cells (DCs), 51
    macrophages, 48–50
    natural killer (NK) cells, 46–47, 46*f*
    type 2 innate lymphoid cells (ILC2 cells), 47–48
  lineage-specific T-cell function, acquisition of, 123–124
Epigenetics
  in adaptive immune system, 51–63
    B-cell development and differentiation, 52–55, 53*f*
    T cells, epigenetic regulation in, 55–60
  definition of, 9–10, 141–142
  immune system and memory in plants, 32–33
  microbiota, 141–142
  plant insect interactions, 30–32, 31*t*
  plant microbe interactions, 29–30
  in transcriptional regulation, 39
Epigenome, 60–61, 150–151
Epigenomics, 161–162, 197*f*
Epistasis, 34
Etanercept, 255
Etoposide, 329–340*t*
Euchromatin, 10
Eukaryotic genome, 40–41
Exemestane, 329–340*t*
Experimental autoimmune encephalomyelitis (EAE) induction, 83–84
Expression quantitative trait DNA methylation (eQTM) analysis, 279
Ezh2-deficient Th2 cells, 86
EZH2 inhibitors, 318

F
FAIRE-seq, 194
Familial cold autoinflammatory syndrome (FCAS), 259–260
Familial Mediterranean fever (FMF), epigenetic deregulation, 258–259
FastQC, 173
Fibroblast-like synoviocytes (FLS), 262–263
Field-programmable gate arrays (FPGAs), 220
Firmicutes, 142–144
Flow cell preparation, Illumina sequencing, 165–167, 166*f*
Fluorescent-activated cell sorting (FACS), 190–191
5-fluorouracil (5-FU), 344
Folate, 151–152
Folic acid, 151–152
Foxo1, B-cell development, 53–54
Foxp3, 78–79, 142–144, 320–321
Free fatty acid receptor 2 *(Ffar2)*, 142–144
Fully methylated regions (FMRs), 176

G

GATA3, 48, 59, 102, 123–124
g-chromVAR, 204
Gemcitabine, 329–340*t*
Gene-based/vertical disease resistance, 32–33
Gene expression regulation, 105–106
Gene regulation by long-distance interactions, 60–63
Genome and transcriptome sequencing (G&T-seq), 196
Genome architecture
  at primary structure scale, 10–11
  at secondary structure scale, 11–13, 12*f*
  at tertiary structure scale, 13–14
Genome Bisulfite Sequencing Analyzer (GBSA), 175
Genome-wide association studies (GWAS), 204, 207–208, 269–270
Germ-free (GF) animals, 139–144, 147–148
Germinal center (GC) B cells, 88, 103
Glesatinib, 329–340*t*
Global Run-On and sequencing (GRO-seq), 226
Glucocorticoid receptors (GR), 18–19
GPR109A, 142–144
G-protein-coupled receptor 43 (GPR43), 142–144
Graphics processing units (GPUs), 220
Graves' disease (GD), 260–261
Guadecitabine, 327–340*t*
Gut microbiome and epigenetics of immune cells, 142–148
  innate lymphoid cells (ILCs), 147
  invariant natural killer T (iNKT) cells, 147–148
  of mononuclear phagocytes, 144–146
  regulatory T cell (Treg), 142–144, 143*f*

H

Hematopoiesis, 187–188
  cancer classification
    biomarkers, 304–305
    computational models, 306–307
    DNA methylation, 305–306
  cell populations, 288
  chronic lymphocytic leukemia (CLL), cell-of-origin
    cancer-specific DNA methylation, 303, 303*f*
    deregulated genes, 301–302
    malignant transformation, 300–301
    methylomics, 302–303
  cis-regulatory elements, 287
  differentiation, 288–290
  epigenetic drugs, 307
    inhibitors of, 307–309
    mechanisms, 309–310
  epigenetic modifications, 286–287
  epigenetic regulation, 45
  hierarchical model of, 287–288, 289*f*
  multipotent progenitor cells (MPPs), 287–288
  myeloid neoplasms
    acute myeloid leukemia (AML), 293–294
    chromosomal aberrations, 294
    complex aberrant karyotype AML (ck-AML), 296
    deletions/monosomies, 295
    DNA methylation, 294
    GATA2 enhancer, 296
    genomic amplifications, 296
    mixed-lineage leukemia (MLL), 294–295
    myelodysplastic syndromes (MDS), 295
  preleukemic mutations, 293
    aging-related stochastic processes, 291
    ASXL1, 293
    DNMT3A, 291–292
    IDH1 and IDH2, 292
    malignancies, 290–291
    TET2, 292
  single-cell epigenomics technologies, 203–204
  T-cell acute lymphoblastic leukemia (T-ALL)
    epigenetic modifiers, 298
    leukemogenesis, 297
    methylation profiles, 298
    oncogenic neo-enhancers, 298–300, 300*f*
  tree-like bifurcation model, 287–288
Hematopoietic progenitor cells (HPCs), 47
Hematopoietic stemand progenitor cells (HSPCs), 288–290
Hematopoietic stem cell (HSC), 2, 3*f*, 48–49, 78, 203–204, 287–288
  B and T cell development, 79*f*
Heterochromatin, 10, 13–14
Heterochromatin protein 1 (HP1), 121
Heterogeneity, 178, 260–261
Heterosis, 34–35
Hidden Markov Model (HMM), 219–221
High-programmed chronic lymphocytic leukemia (HP-CLL), 302–303
High-throughput chromosome conformation capture followed by sequencing (HiC-seq), 195
HiSeq, 164

Histone acceptors, 16
Histone acetylation, 42–43, 47, 84–85, 141–144
Histone acetylase transferase (HAT) inhibitors, 317
Histone acetyl transferase (HAT) activity, 42–43, 256
Histone chaperones, 16
Histone code hypothesis, 17, 162
Histone deacetylase (HDAC) 7, 87–88
Histone deacetylase inhibitors (HDACi), 142–145, 143*f*, 150–151, 262–263, 316–317, 321–323, 342–344
  antigen-presenting cells (APCs), 323
  B cells, 322
  clinical significance of, 325–341
  combination therapy, and DNMTi, 344–345
  dendritic cells (DCs), 323
  immunomodulatory effects, 321–323
  macrophage, 323
  myeloid-derived suppressor cells (MDSCs), 322
  radiotherapy, 344
  regulatory T (Treg) cells, 322
  T cells, 321–322
Histone H3 lysine K4 trimethylation (H3K4me3), 121, 124–126, 145–147, 150
Histone H3 lysine K27 trimethylation (H3K27me3), 85–86, 121, 123–128, 130–132, 141, 145–146, 162, 320
Histone methylation
  immune memory cell differentiation and function, 120–121
  regulating genes in different cellular functions, 126*f*
Histone methyltransferases (HMTs), 130, 131*f*, 243, 341
Histone modifications, 14–15, 42–43, 253–254
  biotic stresses, 29
  epigenetic modifications, 42–43
  interdependence of DNA and, 43
  key epigenetic drivers, 42–43
  in lymphocyte differentiation, 84–86
    histone acetylation, 84–85
    H3K27me3, 85
    H3K4 methylation, 84–85
    writers, readers, and erasers in, 86–88
  microbiota, 141
  sequencing technologies for epigenetics, 171
Histone-modifying enzymes, 17
Histone proteins, 120–121
Histones, posttranslational modifications, 16–18
Histone turnover, 16–17
H3K4me1, 87
H3K4me3, 121, 124–126
H3K9me3, 130–132
H3K4 methylation, 84–85, 257–258
Human hepatocellular carcinoma (HCC) tumor tissue, 205–206
2-Hydroxyglutarate (2HG), 292
5-hydroxymethylcytosine (5hmC), 83–84, 170, 192–193
Hygiene hypothesis, 152
Hypomethylation, 257

I

I-BET762, 317–318
iChIP, 207
Illumina Genome Analyzer, 164
Illumina sequencing, 164–165
  flow cell preparation, 165–167, 166*f*
  library preparation, 164–165, 165*f*
  sequencing by synthesis, 167–169, 168*f*
Immature dendritic cells (DCs), 2, 3*f*
Immediate early genes, 14–15
Immune cell
  cell-type-specific features, 243
  deconvolution methods, 244, 245*t*
  development, 2–5, 3*f*
  DNA methylation (DNAm)
    CD8 T cells, 241
    macrophage differentiation and functions, 240–241
    methylated CpG, 239–240
    myeloid lineages, 240
    self-tolerance and tissue homeostasis, 241
    tumor microenvironment, 241–242
  histone modifications, 243
  perspectives, 248–250
  reference-based models, 244–247, 246*f*
  reference-free models, 247–248
  RNA modification, 242
  T cells
    B cells development, 3–4
    innate lymphoid cells (ILCs), 4–5
    natural killer (NK) cells, 4
    T cells development, 2–3
  technologies, 243
Immune memory cell differentiation and function
  active regulation of chromatin state, 129–132, 131*f*
  CD8+ T-cell-specific transcription factors, 127–129

Immune memory cell differentiation and function *(Continued)*
  DNA methylation, in regulating gene transcription, 122–123
  epigenetic mapping, T-cell memory, 126–127
  epigenetic regulation
    CD8 T-cell function, acquisition of, 124–126, 125*f*
    lineage-specific T-cell function, acquisition of, 123–124
  histone methylation and pattern, acquisition of function, 120–121
Immune response, 5–7, 7*f*
Immune surveillance, 1–2
Immune system development, 1–2
  adaptive immune system, epigenetics in, 51–63
    B-cell development and differentiation, 52–55, 53*f*
    T cells, epigenetic regulation in, 55–60
  epigenetic modifications
    accessible chromatin, 44
    chromatin remodeling and noncoding RNA (ncRNA), 44
    DNA methylation, 41–42
    histone modifications, 42–43
    priming gene for activation, 44–45
    transcriptional control of, 41
  epigenetic regulation, 45–63
    dendritic cells (DCs), 51
    hematopoiesis, epigenetic regulation, 45
    innate immune cells, development of, 45–51
    macrophages, 48–50
    natural killer (NK) cells, 46–47, 46*f*
    type 2 innate lymphoid cells (ILC2 cells), 47–48
  nonspecific (innate) branches, 1–2
  specific (adaptive) branches, 1–2
Immunodysregulation, polyendocrinopathy, enteropathy, X-linked (IPEX) syndrome, 2–3
Immunomodulatory effects, 319–325
  DNA methyltransferase (DNMT) inhibitors on immune cells, 324–325
  HDAC inhibitors (HDACi) on immune cells, 321–323
Immunotherapy, 342–343
Independent surrogate variable analysis (ISVA), 247
Index sorting, 190–191
Indoleamine 2,3-dioxygenase (IDO), 255
Induced regulatory T cells (iTregs), 58, 83–86
Induced systemic resistance (ISR), 29–30
Inflammatory rheumatic diseases, epigenetic biomarkers in, 261*t*
Innate immune cells, 1–2, 45–51
  dendritic cells (DCs), 51
  macrophages, 48–50
  memory in, 9*f*
  natural killer (NK) cells, 46–47, 46*f*
  type 2 innate lymphoid cells (ILC2 cells), 47–48
Innate immunity, plant, 32–33
Innate lymphoid cells (ILCs), 4–5, 147
Inner-City Asthma Consortium (ICAC) cohort, 274
Insulators/boundary elements, 10–11
  properties, 10–11
Intelligent priming systems, 32–33
Intelligent strategies, 30–31
Interferon-γ (IFN-γ), 4–5, 47, 60–62, 85–86, 91, 320
Interferon regulatory factor (IRF) signaling, 323
Interleukins-1β (IL-1β), 259–260
Interleukins-2 (IL-2), 321–322
Interleukins-17 (IL-17), 256–257
Interleukins-18 (IL-18), 259–260
Intermediate-programmed chronic lymphocytic leukemia (IP-CLL), 302–303
Intestinal flora, 139–142
Invariant natural killer T (iNKT) cells, 147–148
Ipilimumab, 329–340*t*

J

Jasmonic acid (JA), 29–33
Jumonji domain containing-3 (JMJD3), 49
Junk DNA, 122–123
Juvenile idiopathic arthritis (sJIA), 262–263

K

Keratinocytes, 256–257
Killer-cell Immunoglobulin-like Receptors (KIR), 46–47
Killer inhibitory receptors (KIR), 4

L

Laser capture microdissection (LCM), 191
Late response genes, 14–15
Leukemia, single-cell epigenomics technologies, 205–206

Leukemic blasts, 293
Levocarnitine, 329–340*t*
Library preparation, Illumina sequencing, 164–165, 165*f*
Life Technology's SOLiD system, 164
Lineage determining factors, 45–46
Lineage-specific T-cell function, acquisition of, 123–124
Lirilumab, 329–340*t*
Locus control regions (LCRs), 10–11
Long-lived memory, 8
Long-lived plasma cells, 8
Long noncoding RNAs (lncRNA), 15–16, 15*t*, 44, 78, 89–90, 100–104
  in B-cell differentiation, 103–104
  in T-cell differentiation, 101–103
Long-range chromatin interactions, 141
  and gene regulation, 13–14
Long-range DNA looping, transcriptional control by, 61*f*
Long-read sequencing technologies, 169
Low-methylated regions (LMRs), 176
Low-programmed chronic lymphocytic leukemia (LP-CLL), 302–303
Lymphocyte differentiation, 77–78, 79*f*
  chromatin-based epigenetic mechanisms in, 80–89
    chromatin accessibility in immune cells differentiation, 88–89
    DNA methylation and lineage commitment, 80–81, 80*f*
    histone modifications, 84–86
  RNA-based mechanisms of regulation of, 89–105
    long noncoding RNAs (lncRNA), 100–104
    microRNA (miRNAs)-mediated regulation of T and B cell differentiation, 90–91
    noncoding RNAs, cross talk between, 104–105
    specific microRNA (miRNAs) regulating T- and B-cell differentiation, 92–100
Lymphocyte induced maturation protein-1 (BLIMP-1), 54–55
Lysine methyltransferases (KMTs), 43
Lysine-specific demethylase 1 (LSD1) inhibitors, 319

M

Machine learning (ML)
  classical classification methods
    Bayesian networks, 227–229
    deep learning, 231–233
    DNA methylation, 230–231
    examples, 226–227
    histone modifications, 230–231
    multiple kernel learning (MKL), 229–230
  classification system, 218–219
  epigenetic code, 217–218
  supervised approaches, 220–221
    area under the receiver operating characteristics (AUROC/AUC), 224–225
    binary classifier, 224
    test dataset, 223–224, 224*f*
    training dataset, 223
  training data generation, 225–226
  unsupervised learning
    ChromHMM, 221
    neural network models, 220
    segmentation methods, 221–223, 222*f*
    training set, 219–220
Macrophages, 8–9, 144–145, 146*f*, 150–151
  lineage commitment of, 50*f*
  M1 subtypes, 49
  M2 subtypes, 49
MACS2, 199
Maintenance methyltransferase, 324
MALAT1, 102
Mature dendritic cells (DCs), 5–6
5mC. *See* 5-methylcytosine (5mC)
MEFV gene, 258–259
Memory, 8–9, 9*f*
  in adaptive and innate immune system, 9*f*
  cells, 60
  definition, 32
  long-lived memory, 8
  in plants, 32–33
Memory T cells, 119
  epigenetic mapping, 126–127
  functional heterogeneity within, 119–120
Metagenomic, 141
Methotrexate (MTX), 255
Methylated CpGs, 255
Methylated DNA immunoprecipitation-seq (MeDIP), 63
Methylation-sensitive amplified length polymorphism (MSAP), 30–31
MethylC-seq, 170
5-methylcytosine (5mC), 162, 170, 192–193
Methyltransferase 1 (MET1), 28
MethyQA, 173

Microbiome and epigenetics of nonmucosal immune cell, 149–152
  epigenetic imprint of microbes on offspring's immune cells, 152
  folic acid, 151–152
  nonmucosal myeloid cells, 150
  short-chain fatty acids (SCFA), 150–151
Microbiota
  composition and numbers of healthy human, 140*f*
  epigenetic mechanism, 141–142
  gut microbiome and epigenetics of immune cells, 142–148
    innate lymphoid cells (ILCs), 147
    invariant natural killer T (iNKT) cells, 147–148
    of mononuclear phagocytes, 144–146
    regulatory T cell (Treg), 142–144, 143*f*
  microbiome and epigenetics of nonmucosal immune cell, 149–152
    epigenetic imprint of microbes on offspring's immune cells, 152
    folic acid, 151–152
    nonmucosal myeloid cells, 150
    short-chain fatty acids (SCFA), 150–151
  skin microbiota and epigenetics of immune cell, 148–149, 149*f*
Micrococcal nuclease (MNase) digestion, 172
MicroRNAs (miRNAs), 54, 78, 89–90
  B-cell differentiation, regulation of, 92–100
  T-cell differentiation, regulation of, 92–100, 93–98*t*
MicroRNAs (miRNAs)-mediated regulation of T and B cell differentiation, 90–91
Midline carcinoma, 317–318
Midostaurin, 329–340*t*
miR-29, 92
miR-10a, 92–99
miR-125a, 92–99
miR-150, 54, 99
miR-181a, 99
Mixed-lineage leukemia (MLL), 294–295, 306, 309, 319, 342
M1 macrophages, 49, 325
M2 macrophages, 49, 325
Mocetinostat, 329–340*t*
Model-based Analysis of ChIP-Seq data (MACS2), 176
Monoclonal antibodies, 342–343
Monocytes, 8–9
Monogenic diseases, 258
Mononuclear phagocytes, 144–146
Mouse cytomegalovirus (MCMV) infection, 150
MPT0G009, 262–263
MS-275, 262–263
Muckle-Wells syndrome, 259–260
Multidimensional scaling (MDS), 219–220
Multi-omics, 19–20
Multi-Omics Factorv Analysis (MOFA), 202
Multiple myeloma (MM), 255–256, 308
Multiplexed single-cell reduced representation bisulfite sequencing (MscRRBS), 192, 205–206
Multipotent precursor (MPP), 51
Multipotent progenitor cells (MPPs), 78, 79*f*, 287–288
MUltiScale enrIchment Calling for ChIP-Seq (MUSIC), 176
Multivariate linear regression, 244–247
Mutant IDH inhibitors (mIDHi), 307–308
Myelodysplastic syndromes (MDS), 295

N

NALP3 protein, 259–260
Nanopore sequencing, 169
Natural killer (NK) cells, 4, 8–9, 88–89, 322
  innate immune cells, development of, 46–47, 46*f*
Natural killer group 2, member D (NKG2D), 322
Natural killer group 2 member D (NKG2D) receptor, 4, 322
Natural killer T (NKT) cells, 93–98*t*
Neonatal onset multisystem inflammatory disease (NOMID), 259–260
Neural network (NN) approach, 231, 233
Next-generation sequencing, 163–164, 177
  disadvantage of, 168
  for DNA methylation, 170
  raw data and quality control, 172–173
  read alignment, 173–174
Next-generation sequencing (NGS)-based single-cell epigenomics methods, 188
Nfil3, 47
Niacin, 142–144
Nicotinamide, 327–328*t*
Nivolumab, 329–340*t*, 342–343
NKG2C, 8–9
NKG2D, 47
N6-methyladenosine (m^{6}A), 242
NOMe-seq, 196

Noncoding regulatory elements work via chromatin folding, 122–123
Noncoding RNA (ncRNA), 44, 78, 104–105, 141
Nonhistone proteins, 18–19
Non-Mendelian epigenetic inheritance, 10, 33
Nonmucosal myeloid cells, 150
Nonnegative least-squares regression (NNLS), 249
Nonnegative matrix fraction (NMF), 247
Nonredundant fraction (NRF), 177
Nonspecific (innate) immune system, 1–2
Notch signaling, 47, 55–57
NovaSeq platforms, 164
Nuclear factor of activated T cells (NFAT), 92
Nucleolus-associated domains, 13–14
Nucleosome editing, 44
Nucleosome occupancy and methylome sequencing (NOMe-seq/scNOMe-seq), 195
Nucleosome positioning, Trans-regulation, 11–13, 12*f*
Nucleosome remodeling, 44
Nucleosomes, 40–41, 84, 120–121, 193–194

O

Obinutuzumab, 329–340*t*
One-drug-fits-all model, 263
Organogenesis, 208
Organ-specific autoimmune disorders, 254
Ovalbumin (OVA), 147–148
Overdominance hypothesis, 34
Oxaliplatin, 329–340*t*
Oxford Nanopore Technologies (ONT), 169
Oxidative bisulfite sequencing (oxBS-seq), 170

P

Paclitaxel, 329–340*t*
Paired box 5 (PAX5), 52–53, 86–87, 103
Paired box 5 (PAX5)-activated intergenic repeat (PAIR) elements, 103–104
Panobinostat, 262–263, 327–340*t*, 342–343
Parental imprinting, 33
Partial correlations (LR), 244–247
Partially methylated domains (PMDs), 176
Pathogen-associated molecular pattern (PAMP)-triggered immunity (PTI), 32–33
Pattern recognition receptors (PRRs), 5–6, 8–9
Peak calling, ChIP-seq data analysis, 176
Pembrolizumab, 329–340*t*
Perforins, 47, 118–119
Pericentromere-associated domains, 13–14
Peripheral blood mononuclear cells (PBMCs), 206, 255–258
Peripheral differentiation of B cells, 54–55
Pertussis toxin (PT), 150–151
Phase errors, 172–173
Phase separation model for transcriptional control, 13–14
Phosphatase and tensin homolog (PTEN), 92
Pinometostat, 342
PIQA, 173
Plant DNA methylation, biotic stresses, 26–27, 27*f*
  de novo DNA methylation, 27–28
  maintenance, 28–29
Plant epigenetics, biotic stresses, 33
  heterosis, 34–35
  somatic embryogenesis, 34
Plant immune system, 32–33
Plant insect interactions, epigenetics of, 30–32, 31*t*
Plant microbe interactions, epigenetics of, 29–30
Plant neurobiology, 32
Plant-pathogen interaction, 29–30
Plant-plant communication, 29–30
Plasmacytoid dendritic cells (pDCs), 51
Poised genes, 44–45, 85, 141
Pol II, 13–15
Pol IV, 28
Pol V, 28
Polycomb repression complex 2 (PRC2), 86, 105, 129, 298
Polyploidy, 35
Post-bisulfite adapter-tagging (PBAT) approach, 191–192
Posttranscriptional gene silencing (PTGS), 29
Posttranscriptional regulation, of RNA expression, 14–16
Posttranslational modifications (PTM), 40–41, 120–121, 122*f*
  chromatin remodeling complexes and histones, 16–18
  nonhistone proteins, 18–19
Posttraumatic stress disorder, 175
Potent cytotoxic cells, 4
Pre-B to immature B cells, 54
Prednisone, 329–340*t*
Prevention and Incidence of Asthma and Mite Allergy (PIAMA), 274–275

Primary and secondary pathogen-specific T-cell responses, 118*f*
Primary-miRNA (pri-miRNA), 90
Priming, 32–33
Principal component analysis (PCA), 200, 219–220, 247–248
Probiotics, 152–153
Pro-B to pre-B cell commitment, 54
Programmed cell death protein-1 (PD-1), 342–343
Promoter regions, 10–11, 14–15
Propionate, 142–144
Propionibacterium acnes, 149
Protein-coding RNA, 15–16
Protein-DNA interaction, 193
Pseudo-bulk analysis, 199
Psoriasis, 261*t*
  epigenetic deregulation, 256–258
Pyrin, 258–259

Q

QC tools, raw data and quality control, 173
5q – syndrome, 295
Quisinostat, 327–328*t*

R

Radiotherapy, 344
Random forest (RF) classifiers, 230
Reactive oxygen species, 29–30
Recombination activating gene (RAG) 2, 54
Recurrent neural networks (rNN), 231–232
Reduced representation bisulfite sequencing (RRBS), 11, 170
Regulatory loops, 13–14
Regulatory noncoding RNA, 15–16, 15*t*
Regulatory T cells (Tregs), 2–3, 55, 58, 63, 92–99, 93–98*t*, 142–144, 143*f*, 255–257
Repressive code, 49
Resident memory T cells (T_{RM}), 120
Resminostat, 327–328*t*
Rheumatic diseases, therapeutic treatments, 262*t*
Rheumatoid arthritis (RA), 254–256
  pathological development of, 254–255
  peripheral blood mononuclear cells (PBMCs), 255–258
  rheumatoid factor (RF), 254–256
Rheumatoid factor (RF), 254–256, 260–261, 261–262*t*
Rituximab, 329–340*t*
RMAP-bs, bisulfite sequencing read aligners, 174
RNA-associated silencing, biotic stresses, 29
RNA-based mechanisms of regulation of lymphocyte differentiation, 89–105
  long noncoding RNAs (lncRNA), 100–104
  microRNA (miRNAs)-mediated regulation of T and B cell differentiation, 90–91
  noncoding RNAs, cross talk between, 104–105
  specific microRNA (miRNAs) regulating T- and B-cell differentiation, 92–100
RNA component of the mitochondrial-RNA-processing endoRNase (Rmrp), 102
RNA-directed DNA methylation pathway (RdDM), 27–28
RNA expression, posttranscriptional regulation of, 14–16
RNA-induced silencing complex (RISC), 29, 90
RNA interference (RNAi), 27–29, 33, 62, 150–151
RNA polymerase II, 13–14
RNA-seq, 177–178
Robust multivariate linear regression, or robust partial correlations (RLR/RPC), 244–247
Roche's 454 sequencing system, 164
Romidepsin, 327–340*t*
Runt domain, 58
RUNX3, 128–129
RUNX3 deficiency, 129

S

Salicylic acid (SA), 29–30
Sanger sequencing, 163, 172–173
ScAba-seq, 192–193
ScAsAT, 200
scHiC-seq by using combinatorial indexing (sciHiC-seq), 195
scipATAC-seq, 194
SCRAT, 200–201
scTHS-seq method, 194–195
scTrio-seq, 205–206
Seed sequence, 90
Sequence Alignment/Map (SAM) format, 174
Sequencing technologies for epigenetics
  applications of, 169–172
    DNA methylation, 170
    histone modifications, 171
  data processing and computational analysis, 172–178
    methylation data analysis, 174–176

- raw data and quality control, 172–173
- read alignment, 173–174
- Illumina sequencing protocol, 164
  - flow cell preparation, 165–167, 166*f*
  - library preparation, 164–165, 165*f*
  - sequencing by synthesis, 167–169, 168*f*
- next-generation sequencing, 163–164
- third-generation sequencing, 168–169
  - nanopore sequencing, 169
  - single-molecule, real-time (SMRT) sequencing, 169

SeqWell, 191
Short-chain fatty acids (SCFAs), 139–144, 143*f*, 146*f*, 149–151
Signal transducer and activator of transcription 1 (STAT1), 83
Signal transducer and activator of transcription 3 (STAT3), 51, 83
Silencers, 10–11
Single-cell assay for transposase-accessible chromatin (scATAC-seq), 88–89, 198, 209
Single-cell DNA adenine methyltransferase identification sequencing (scDamID-seq), 193
Single-cell epigenomics technologies, 19–20, 185–188, 186–187*t*, 189–190*t*
- chromatin structure, 193–196
  - chromosome conformation, 195
  - single-cell multiomics, 195–196
- computational challenges and solutions, 196–202, 197*f*
  - downstream analysis, 200–202
  - epigenomics techniques and generic analysis pipeline, 197*f*
  - quality control and preprocessing, 198–200
- DNA modifications, 192–193
- evolution of, 188*f*
- fluorescent-activated cell sorting (FACS), 190–191
- immune system, studying of, 202–208
  - aging, 206–208
  - hematopoiesis, 203–204
  - leukemia, 205–206
- laser capture microdissection (LCM), 191
- protein-DNA interaction, 193
- with spatial resolution, 208

Single-cell hi-C protocols, 200
Single-cell level using reduced-representation bisulfite sequencing approach (scRRBS-seq), 192
Single-cell methylome and transcriptome sequencing technique (scM&Tseq), 196
Single-cell multiomics, 195–196
Single-cell nucleosome, methylation, and transcription sequencing (scNMT-seq), 196, 202–203
Single-cell omics technologies, 191–192, 195–197, 205–206
Single-cell RNA-seq, 129, 202–203
Single combinatorial indexing ATAC followed by sequencing (sciATAC-seq), 194, 207–208
Single-molecule, real-time (SMRT) sequencing, 169
Single nucleotide polymorphic (SNP), 233
Sitravatinib, 329–340*t*
Skin microbiota and epigenetics of immune cell, 148–149, 149*f*
Small interfering RNAs (siRNAs), 29
SnapATAC, 200
Somaclonal variation, 34
Somatic embryogenesis, 34
Somatic hypermutation (SHM), 78
Spatial resolution, single-cell epigenomics technologies with, 208
Special AT-rich Binding Protein 1 (SATB1), 57–58, 62–63
Specific (adaptive) immune system, 1–2
Staphylococcus epidermidis, 148–149
STAR, 173–174
STAT3. *See* Signal transducer and activator of transcription 3 (STAT3)
Statistical learning theory, 219
Stem cell-like memory T cells (T_{SCM}), 119–120
STREAM, 201
String-of-beads structure, 11
Structural loops, 13–14
Sub-topologically associated domains, 13–14
Supervised learning, 220–221
- area under the receiver operating characteristics (AUROC/AUC), 224–225
- binary classifier, 224
- test dataset, 223–224, 224*f*
- training dataset, 223

Support Vector Regressions (SVR), 244–247
Surrogate variable analysis (SVA), 247
SUV39H1, 130
SWItch Sucrose nonfermentable (SWI/SNF) complex remodelers, 44
Systemic acquired resistance (SAR), 29–30

Systemic autoimmune rheumatic diseases (SARDs), 254
Systemic lupus erythematosus (SLE), 254, 262–263
Systemic resistance, 29–30
Systems immunology, 19–20

T

Tamoxifen, 329–340*t*
Tazemetostat, 327–328*t*
T-cell acute lymphoblastic leukemia (T-ALL)
  epigenetic modifiers, 298
  leukemogenesis, 297
  methylation profiles, 298
  oncogenic neo-enhancers, 298–300, 300*f*
T-cell differentiation
  long noncoding RNAs (lncRNA), 101–103
  microRNA (miRNAs)-mediated regulation of, 90–91
  specific microRNA (miRNAs) regulating, 92–100
T-cell receptor (TCR), 1–4, 81
T cell receptor (TCR) beta, 55–57
T cells, 78, 82–83
  bivalent marking in, 85–86
  CD4+ T cells, 117–119
  CD8+ T cells, 117, 119
  development, 2–3, 78–79
  epigenetic regulation in, 55–60
    CD4/CD8 lineage commitment, 57–58
    common lymphoid progenitors (CLPs) to T cells, 55–57, 56*f*
    gene regulation by long-distance interactions, 60–63
    plasticity and memory, epigenetic changes in, 59–60
    terminal differentiation and function, in periphery, 58–59
TCF1, 48, 57
Tefinostat, 327–328*t*
t-distributed stochastic neighbor embedding (t-SNE), 201
Tefinostat, 327–328*t*
Temozolomide, 344
Tet2, 83–84
Tet-assisted bisulfite sequencing (TAB-seq), 170
Tetrahydrouridine, 329–340*t*
TFBSs, 128
T helper type 1 (Th1) cells, 6, 49, 59, 62, 93–98*t*, 320
T helper type 2 (Th2) cells, 6, 49, 86–87, 93–98*t*, 102, 320
T helper type 2 (Th2) cytokines, 62
T helper type 17 (Th17), 93–98*t*, 102, 142, 320–321
Third-generation sequencing, 163, 168–169, 172, 178
  DNA methylation, 170
  nanopore sequencing, 169
  read alignment, 174
  single-molecule, real-time (SMRT) sequencing, 169
ThPOK, 57–58
Thymic stromal lymphopoietin (TSLP) cytokines, 47
Tissue homeostasis, macrophages, 48
Tissue-resident memory T cells (T_{RM}), 8, 9*f*
Toll-like receptors (TLR), 5–6, 49
Topologically associating domains (TADs), 13–14, 286–287
Trained immunity, 145–146
Transcriptional enhancers (TEs), 122–123
Transcriptional regulation, epigenetics in, 39
Transcription factors (TFs), 118–119, 321
  CD8+ T-cell, in chromatin remodeling and acquisition of function, 127–129
  common lymphoid progenitors (CLPs) to T cells, regulation by, 55–57, 56*f*
Transcription start sites (TSSs), 200
Transgenerational memory, 30–31
Transposase-accessible chromatin with sequencing (ATAC-seq), 250
Transposome hypersensitive sites sequencing (THS-seq) assay, 194–195
Trans-regulation, chromatin packaging and nucleosome positioning, 11–13, 12*f*
Tranylcypromine, 327–328*t*
Trastuzumab, 329–340*t*
Treg cells. *See* Regulatory T cells (Tregs)
Treg-specific demethylation region (TSDR), 2–3, 241
Tremelimumab, 329–340*t*
Trichostatin A (TSA), 262–263
Trimmomatic, 198

Trypsin protease inhibitors (TPIs), 33
Tumor-associated macrophages, 325
Type 2 innate lymphoid cells (ILC2 cells), 47–48

U
Ubiquitin-like, with pleckstrin-homology and RING-finger domains 1 gene (Uhrf1), 144
Uniform Manifold Approximation and Projection (UMAP), 201
Universal amplifier genes, 14–15
Unmethylated regions (UMRs), 176
Unsupervised learning
  ChromHMM, 221
  neural network models, 220
  segmentation methods, 221–223, 222*f*
  training set, 219–220
Utomilumab, 329–340*t*

V
Valproic acid, 329–340*t*
Varinostat, 262–263
Viral infection, 128–129
Vitamin B_9. *See* Folic acid
Vorinostat, 327–340*t*

W
Web service for Bisulfite Sequencing data Analysis (WBSA), 175–176
Whole genome bisulfite DNA sequencing (WGBS), 11, 170, 203–204, 206
Whole-genome sequencing, 168
Wound healing process, 148

Z
ZFP57, 151

This page intentionally left blank

Epigenetics of the Immune System

Edited by
Dieter Kabelitz and Jaydeep Bhat

Immunology is in the heart of "translational epigenetics" because an ever-increasing range of novel therapies are based on epigenetic modifications of the immune system.

Epigenetics of the Immune System focuses on different aspects of epigenetics and immunology. Readers will learn about the fundamental mechanisms relating to epigenetics and immune system. Next an insight into the epigenetics as a toolbox for exploring processes involved in the immune cell development, differentiation, and functional responses is provided; also in-depth information on immune cell regulation and mechanisms using "omics" and computational biology approaches is given. Finally, Epigenetics of the Immune System also illustrates the translational importance of epigenetics and the immune system for understanding pathophysiology of diseases and its therapeutic applications.

Key Features

- Provides the combination of an overview of most important immune mechanisms, current status of epigenetics, and how both these are brought together to more precisely understand the immune system and its epigenetic regulation.
- Presents key principles of immune mechanisms in epigenetics.
- Features in-depth chapter contributions from a wide range of international researchers and specialists in immunology, translational medicine, and epigenetics.
- Merges two large areas and covers the unique interrelatedness of epigenetics and immunology.

ELSEVIER

ACADEMIC PRESS
An imprint of Elsevier
elsevier.com/books-and-journals

ISBN 978-0-12-817964-2
9 780128 179642